DOCUMENT RESUME

ED 099 817 CS 001 538

AUTHOR	Siegel, Arthur I.; Burkett, James R.
TITLE	Application of Structure-of-Intellect and Psycholinguistic Concepts to Reading Comprehensibility Measurement.
INSTITUTION	Air Force Human Resources Lab., Lowry AFB, Colo. Technical Training Div.
REPORT NO	AFHRL-TR-74-49
PUB DATE	Sep 74
NOTE	160p.
EDRS PRICE	MF-$0.75 HC-$7.80 PLUS POSTAGE
DESCRIPTORS	Communication (Thought Transfer); *Computational Linguistics; Military Personnel; *Psycholinguistics; Psychometrics; *Readability; Reading Achievement; *Reading Comprehension; *Reading Research

ABSTRACT

Air Force training depends to a considerable extent on the transmission of information through textual materials. This study was performed to define and clarify methods for increasing the capability of written materials to transfer information to the reader. Variables drawn from current psycholinguistic and intellective function literature were defined and related to current theories of learning. Then, for each variable, two sets of reading materials were developed and administered to the airmen. After reading the materials, the airmen were tested to determine how well they comprehend the various materials. Comparison of the comprehension of materials which were heavily loaded with psycholinguistic and intellective variables with comprehensibility of materials which were low on these variables provided the basis for statements relative to the effects of manipulating these variables. The results were recorded and analyzed and new techniques for judging the comprehensibility of textual materials have been developed and made available. (Author/TS)

AFHRL-TR-74-49

AIR FORCE HUMAN RESOURCES LABORATORY

APPLICATION OF STRUCTURE-OF-INTELLECT AND PSYCHOLINGUISTIC CONCEPTS TO READING COMPREHENSIBILITY MEASUREMENT

By

Arthur I. Siegel
Applied Psychological Services, Inc.
Wayne, Pennsylvania 19087

James R. Burkett

TECHNICAL TRAINING DIVISION
Lowry Air Force Base, Colorado 80230

September 1974

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever. and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications. or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report was submitted by Applied Psychological Services, Inc., Wayne, Pennsylvania 19087, under contract F41609-73-C-0018, project 1121, with Technical Training Division, Air Force Human Resources Laboratory (AFSC), Lowry Air Force Base, Colorado 80230. Dr. James R. Burkett, Technical Training Division, was the contract monitor.

This report has been reviewed and cleared for open publication and/or public release by the appropriate Office of Information (OI) in accordance with AFR 190-17 and DoDD 5230.9. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service (NTIS).

This report has been reviewed and is approved.

MARTY R. ROCKWAY, Technical Director
Technical Training Division

Approved for publication.

HAROLD E. FISCHER, Colonel, USAF
Commander

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE — READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFHRL-TR-74-49	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) APPLICATION OF STRUCTURE-OF-INTELLECT AND PSYCHOLINGUISTIC CONCEPTS TO READING COMPREHENSIBILITY MEASUREMENT		5. TYPE OF REPORT & PERIOD COVERED Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Arthur I. Siegel James R. Burkett		8. CONTRACT OR GRANT NUMBER(s) F41609-73-C-0018
9. PERFORMING ORGANIZATION NAME AND ADDRESS Applied Psychological Services, Inc. Wayne, Pennsylvania 19087		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62703F 11210408
11. CONTROLLING OFFICE NAME AND ADDRESS HQ, Air Force Human Resources Laboratory (AFSC) Brooks Air Force Base, Texas 78235		12. REPORT DATE September 1974 13. NUMBER OF PAGES 164
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Technical Training Division Air Force Human Resources Laboratory Lowry Air Force Base, Colorado 80230		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

readability
comprehensibility
computational linguistics
syntactical analysis
text analysis
semantic information processing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Concepts and considerations relative to the measurement of textual readability/comprehensibility through psycholinguistics and through intellective variables are presented. Two experiments, each investigating the potential of each of these areas for readability/comprehensibility measurement, are described and the results of these experiments are given. The results supported a contention favoring the potential of psycholinguistics and intellective concepts for readability/comprehensibility measurement. Concepts are presented for automating the measurements described.

DD FORM 1473, 1 JAN 73 — EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY

Problem

Air Force training depends to a considerable extent on the transmission of information through textual materials. This present work was performed to define and clarify methods for increasing the capability of written materials to transfer information to the reader.

Approach

Variables drawn from current psycholinguistic and intellective function literature were defined and related to current theories of learning. Then, for each variable, two sets of reading materials were developed: (1) a set which is heavily weighted in the variables, and (2) a set which is lightly weighted. These sets of materials were administered to airmen. After reading the materials, the airmen were tested to determine how well they comprehended the various materials. Comparison of the comprehension of the materials which were heavily loaded on the psycholinguistic and the intellective variables with the comprehensibility of the materials which were low on these variables provided the basis for statements relative to the effects of manipulating these variables on comprehensibility.

Results

The effects of psycholinguistic and intellective function related variables on the comprehensibility of written materials were demonstrated. Methods for improving the readability/comprehensibility of textual materials were defined. New techniques for judging the comprehensibility of textual materials have been made available, and initial insights into methods for computer analysis of text have been developed.

Conclusions

Written training materials can now be made more efficient and cost-effective. The findings are not only pertinent to increasing the comprehensibility of training materials but also to the comprehensibility of all written materials.

PREFACE

The editors wish to acknowledge the contribution of a number of individuals at Applied Psychological Services, Inc., in preparing this report. Chapter II, Readability/Comprehensibility as Related to the Structure-of-Intellect Model, was written by Arthur I. Siegel and Brian A. Bergman. Chapter III, Psycholinguistic Determinants of Readability, was authored by Joseph V. Lambert and Arthur I. Siegel. Finally, Chapter IV, Feasibility of Automatic Calculation of Readability/Comprehensibility Metrics, was prepared by J. Jay Wolf and Arthur I. Siegel.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

TABLE OF CONTENTS (cont.)

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF TABLES (cont.)

CHAPTER I

OVERVIEW

For several years, Applied Psychological Services, under the sponsorship of the Air Force Human Resources Laboratory, has been engaged in research into methods for increasing the comprehensibility of written materials as employed in Air Force technical training. Such research will become increasingly important if the literacy level of the typical recruit declines within the all volunteer service concept. Mainly, however, increasing the comprehensibility of the textual materials employed in the training situation clearly can be expected to reduce training time and costs and to increase training effectiveness.

The initial efforts of the program produced a comprehensive review of methods for measuring readability/comprehensibility (Williams, Siegel, & Burkett, in press) along with experimental data relative to the questions of how and in what training context auditory supplementation of written materials will increase the transfer of knowledge (Lautman, Siegel, Williams, & Burkett, in press). During the course of the prior work, it became evident that currently available methods of measuring the readability/comprehensibility of textual materials are less than adequate. Such methods (Flesch, 1943; Dale & Chall, 1948; Smith & Sentner, 1970) rely on frequency of common word use, word length counts, sentence length counts, and the like as a basis for measuring the readability of textual materials. Quite obviously, such counts fail to consider: (1) the familiarity of the reader with the subject content vocabulary, and (2) the inherent mental or intellectual load placed on the reader by the textual materials.

In regard to the first point (familiarity of the reader with the subject content vocabulary), the word "zeitgeist" will be unfamiliar to most laymen but highly familiar to most behavioral scientists. Accordingly, the sentence "His work was not in tune with the zeitgeist" will be highly comprehensible to most behavioral scientists but difficult for others. Accordingly, it seems that word frequency, as found in most lists of familiar words in the English language, cannot be employed as an index of the readability of technical text. Similar arguments may be advanced vis-a-vis the metric basis for other aspects of current readability indices which rely on such counts.

In regard to the second point (conceptual difficulty of the materials), available methods of measuring readability obviously fail to consider such aspects as the memory load placed on the reader, the inductive or deductive reasoning involved in mastering the text, the number of stimulus-response units involved, and the length of the various chains, the clarity of the multiple discriminations involved, and the like. A text which contains the statement "The reader will be able to derive this equation for himself" will be more difficult for most readers than the text which does not place this mental load on the reader. The importance of decreasing the mental load on the reader as a technique for increasing readability/comprehensibility was initially evidenced in a study performed by Siegel and Siegel (1953), who Flesch analyzed the major preelection speeches of Eisenhower and Stevenson. Stevenson had been criticized during the course of the campaign for speaking at too high a level. The Flesch analysis failed to indicate any difference between the speeches of the two candidates. The conclusion to be drawn is that Stevenson's words weren't any bigger or less familiar than Eisenhower's--the problem was in the depth and intellective involvement required by Stevenson's thoughts.

Gestalt and behavior theory principles seem to be especially relevant to arguments favoring a more wholistic analysis of reading difficulty. Gestalt psychology has taught us that learning begins with a whole, not with elemental parts. The whole, in perception and learning, is more than the sum of the parts. In reading, one must attend to certain nonphysical aspects of the reading situation (e.g., relationships, proximity, ambiguity, closure, meaningfulness, context, etc.). Most of the prior elementistic formulations to measuring readability/comprehensibility are unable to account for the additional difficulty caused by variations in the wholistic aspects of textual material. The basis for the current approach is that a more fruitful and diagnostic approach to readability/comprehensibility measurement would include the nonphysical attributes of textual material. Inclusion of the structure-of-intellect and the psycholinguistic involvement imposed by reading material provides an opportunity for a wholistic analysis, since these factors cannot be considered in an elementalistic fashion.

Purpose of Present Program

Accordingly, the present program focuses on the development of techniques which reflect the readability/comprehensibility of textual materials on the basis of the intellective involvement inherent in comprehending the materials. To this end, two separate but related approaches to readability/comprehensibility measurement of textual materials were investigated: (1) an approach which is based on and drawn from the Guilford structure-of-intellect model, and (2) an approach which is based on contemporary psycholinguistics.

The logic, methods, and findings of the approach based on the structure-of-intellect constructs are described in Chapter II, while Chapter III presents a similar description relative to the psycholinguistic approach. Finally, Chapter IV of this report presents a description of techniques for automatically deriving the structure-of-intellect and psycholinguistic metrics (developed and described in Chapters II and III) through digital computer methods.

CHAPTER II

READABILITY/COMPREHENSIBILITY AS RELATED TO THE STRUCTURE-OF-INTELLECT MODEL

The structure-of-intellect (SI) model (Guilford, 1967; Guilford & Hoepfner, 1971) was developed by Guilford in conjunction with his research on intellectual abilities over a 20 year period. Many years of factor analytic research by Guilford and his associates at the University of Southern California produced a hypothetical construct as to the nature and structure of human intellectual activity.

The SI model is a cross classification model that classifies intellectual abilities along three different dimensions. Each dimension is divided into categories which intersect with the categories of the other dimensions of ability. Mental operations represent one dimension of classification in the SI model. The five mental operations are: (a) cognition, (b) memory, (c) divergent production, (d) convergent production, and (e) evaluation.

The second classification dimension of the SI model involves the content areas of information on which the mental operations are performed. These areas of information include: (a) figural, (b) symbolic, (c) semantic, and (d) behavioral. Thirty separate abilities can be derived from the combination (intersection) of the five categories in the mental operation dimension and the four categories in the contents dimension.

The final dimension of intellect in the SI model concerns the formal types of information dealt with. These informational types or products can be units, classes, relations, systems, transformations, and implications. When the six products are combined with the five operations and with the four contents, 120 orthogonal abilities result. The SI model is composed of these 120 abilities. The SI model, then, can be viewed as a three dimensional cube. This cube is shown in Figure 2-1.

BEST COPY AVAILABLE

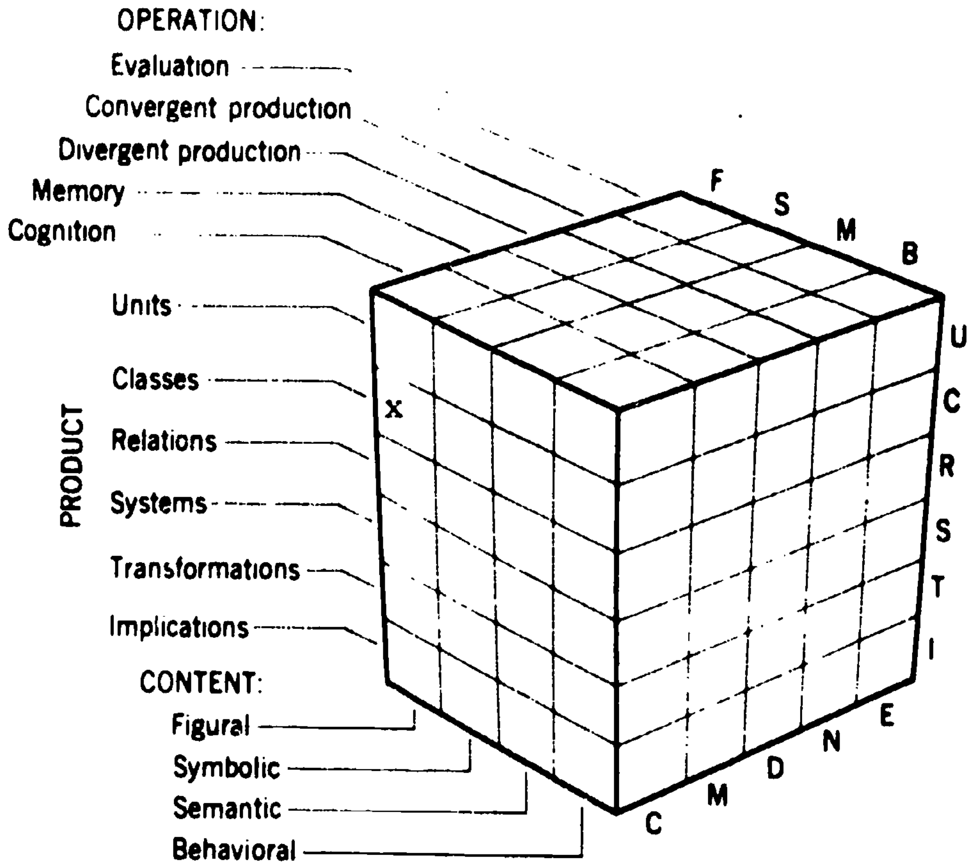

Figure 2-1. The structure-of-intellect model (from Guilford & Hoepfner, 1971).

No prior research has been performed on the adaptation of the SI model to readability/comprehensibility. The reason for this is easy to understand. The SI model is not couched in terms that can be readily related to readability/comprehensibility. Guilford describes his SI model in ability or tested ability terminology. Our first step, then, is to convert Guilford's tested ability concepts into readability/comprehensibility concepts. Specifically, our contention is that textual materials which require high levels of SI abilities for mastery can be said to be less readable/comprehensible than materials which require lower levels of these abilities. The problem then becomes that of deriving metrics which can be applied to textual materials and which reflect the SI abilities required to master the materials. This involves adopting the SI model to a readability/comprehensibility format such that the degree to which a particular reading selection is loaded in various SI factors may be quantified. Since the SI model contains 120 cells (abilities), a sample was required. To this end, those SI abilities which seemed most relevant to the readability/comprehensibility problem in the Air Force technical training context were selected for study. The abilities so selected were: cognition of semantic units, cognition of semantic relations, memory of semantic units, memory of figural units, convergent production of semantic implications, convergent production of semantic systems, divergent production of semantic units, and evaluation of symbolic units. Each of these is expanded on categorically below.

Factors Derived from the SI Model Involved in Readability/Comprehensibility

Cognition of Semantic Units (CMU)

Cognition of semantic units in the readability/comprehensibility context is defined as the extent to which the text forces the reader to recognize a diversity of word forms. Thus, the rhyme "One little piggy went to market, one little piggy stayed home" is held to be readable because of the common word use. The redundancy of words is held to increase readability. The same material written as "A unitary small piggy went to market, one little hog stayed home" is held to be less readable than the original text.

Cognition of Semantic Relations (CMR)

Cognition of semantic relations is defined as the extent to which the text forces the reader to recognize the relationship between two items or words. Guilford (1967) uses analogy and word linkage tests to measure this factor. In word linkage tests, the test taker is required to match sets of words in terms of their relatedness or connectedness. This factor can be varied in reading material by requiring the reader to form analogies or word linkages while reading. One would expect that increasing the requirement for relational thinking in a reading selection would decrease reading comprehension.

Memory of Semantic Units (MMU)

Memory of semantic units is synonymous with memory for meaning and facts. Guilford (1967) uses a memory for ideas test to measure this factor. By implication, it can be held that textual materials with a higher degree of replication of various facts and ideas will be more comprehensible than materials with a lower degree of such repetition.

Evaluation of Symbolic Units (ESU)

In order to comprehend symbolic units, a mental conversion is required. Guilford (1967) used an abbreviation test to measure ability to evaluate symbols. Accordingly, the sentence "The C.I.O. is affiliated with the A.F. of L." is held to be more difficult than the same text in nonabbreviated form. The logic here holds that persons will better remember and comprehend the material when it is expressed in semantic form than when it is expressed in symbolic (abbreviated or acronym) form.

Memory of Figural Units (MFU)

The memory for figural units involves the ability to reconstruct unitary facts presented in figural as opposed to textual form. Guilford and Hoepfner (1971) used map reading tests to measure ability on this factor. In the readability/comprehensibility context, we hold that the figurally presented information which presents the greatest memory load on the reader will be the most difficult.

Divergent Production of Semantic Units (DMU)

According to Guilford (1967), divergent production of semantic units involves the ability to enumerate class members given certain class properties. With regard to the readability of training texts, inclusion of divergent production of semantic units would require the reader to enumerate class members on his own rather than have the class member supplied by the reading selection. We hypothesize that the requirement for divergent production in reading materials will yield decreased readability/comprehensibility.

Convergent Production of Semantic Systems (NMS)

In the measurement of convergent production of semantic systems ability, Guilford (1967) used tests of ordering. To extrapolate to the readability/comprehensibility context, one would conjecture that the readability of organized material is greater than that of the same material with the sentences arranged in a less organized format.

Convergent Production of Semantic Implications (NMI)

Guilford and Hoepfner (1971) use symbolisms, attribute listing, missing links, and sequential association tests to measure convergent production of semantic implications. Reading material loaded in convergent production requires the reader to perform syllogistic reasoning tasks. Material which does not require this ability would complete the syllogism for the reader. Increase of the convergent production of semantic implications load in a text should decrease comprehensibility.

Related Literature

The literature analysis that follows is both theoretical and eclectic in that the ideas of men with widely varying theoretical persuasions are employed to support the selected SI factors as related to textual readability/comprehensibility. The criteria for inclusion in this analysis is that the ideas represented bear some relevance, either by analogy or directly, to the SI factors as readability metrics.

Generally, the bulk of the ideas represented come from four sources: (a) reading literature, (b) classical behavior theory including reinforcement theory, contiguity, etc., (c) gestalt theory, and (d) phenomenological theory.

Both Briggs (1966) and Lumsdaine (1966) feel that the theoretical constructs of learning theory must be accounted for so that instructional materials can be improved. Some of the ideas depicted here have been used to support our selection of all the SI factors collectively; other ideas apply to only one or a few of the selected factors.

Basically, any situation which requires reading comprehension also requires learning. That is, a reading situation is also a learning situation. The reader may not be required to learn the reading passage or book totally, but he is required to remember the concepts, facts, relationships, and implications presented in the reading. According to Gagné (1965), learning is a change in human capacity not dependent upon maturation or growth. "The kind of change called learning exhibits itself as a change in behavior, and the inference of learning is made by comparing what behavior was made possible before the individual was placed in a 'learning situation' and what behavior can be exhibited after such treatment. The change may be, and often is, an increased capability for some type of performance" [p. 5]. The above definition given by Gagné corresponds precisely to the paradigm of an individual prior to and after he reads textual material.

The Reading Literature

The bulk of the reading literature tends to be supportive of the "cognition of semantic units" factor. Essentially, this ability reduces the vocabulary load that the text places upon the reader. As word diversity increases, the potential for unfamiliar or novel words increases. This places a greater vocabulary load on the individual reader.

Gray and Leary (1935) separated good from poor readers and found that different factors accounted for or correlated with comprehension scores. For "poor" readers vocabulary correlated the highest with comprehension, but for "good" readers sentence length and structure correlated the highest with comprehension.

Dale and Tyler (1934) used three principles to produce reading selections which were "easy" to read: (a) use of very basic vocabulary, (b) use of informal style characterized by conversational manner and anecdotal examples, and (c) freedom from digression from the topic of interest. The number of technical words in a passage was found by Dale and Tylor to be a correlate of comprehension. Similarly, George Miller (1951) indicated that short familiar words are easier to read than unfamiliar words. Finally, Lorge (1944) found that vocabulary was the most important single determinant of readability.

Various authors have used different methods to measure the vocabulary difficulty in a reading selection. One of these methods is the proportion of words not appearing on Dale's list of 769 common words (Lorge, 1944; Spache, 1953; Gray & Leary, 1935).

Another measure, the "type/token ratio," is the ratio of different words to the total number of words in a passage. The type/token ratio is an index of communication flexibility or variability (Osgood, 1953).

Flesch (1943) thought that abstractness as well as other variables could be included in a readability formula. Flesch used the syllable count per 100 words as his measure of abstractness. In the context of the present research, we expect that any reading selection that requires evaluation of symbolic units would be more difficult than straight prose. By their very nature, symbols are compact abstract representations or abstractions, (words, thoughts, etc.) and they require an inordinate amount of time to read and remember when compared to normal prose.

One of Harris' (1961) principles for remembering what one has read is that "material that is well organized in the reader's mind is easier to remember than material which is unorganized. The efficient reader tries to grasp the author's plan and to understand the relationships between ideas and the relations between the major ideas and the facts or details which give them definite meaning" (p. 445). The above statement can be applied to several of the selected structure-of-intellect factors. First, when reading material requires cognition of semantic relations or convergent production of semantic systems then, by implication, we can infer that the material is less well organized than it could be. A requirement for relational thinking in reading means that the thoughts and ideas in the passage are not logically, or contiguously related. Accordingly, if a reader must provide relations not provided by the reading passage, he will have a more difficult time comprehending and remembering the selection.

In a like manner, reading material which requires convergent production of semantic systems is relatively disorganized. For example, if a reader is presented with a disorganized system of sentences (parts) and for comprehension to be evidenced the reader must be able to assimilate and integrate these parts into a systematic whole, the passage will be difficult. Assimilation and integration, in this context, is akin to organizing the parts of the systems into a sensible whole so that conceptual understanding can occur.

Harris (1961) also indicates that reading material which is too difficult for a reader can affect his concentration and effort with the resultant loss in comprehension. "The children who can maintain good effort and concentration when working on very difficult material usually do not become remedial problems" (p. 462). Harris' contention possesses implications for all of the selected structure-of-intellect factors. Each factor, when incorporated into reading material, requires extra effort and concentration on the part of the reader.

Classical Behavior Theory

Generally, the ideas presented in the following sections were derived without adherence to any behavioral learning theoretic concept. The intent here is to relate, either by analogy or inference, behavior theoretic concepts to the SI factors in the readability/comprehensibility context.

Distributed Practice vs. Massed Practice. According to Hovland (1951), interference is the factor which dissipates during distributed practice. Thus, distributed practice is preferred to massed practice which produces interference. Accordingly, the more tightly woven a series of facts in a reading selection, the more difficult the readability/comprehensibility of the selection. The more test pauses in the selection--in which new facts can be assimilated by the reader--the easier it will be to read. The foregoing ideas, then, can be used to support the memory for semantic units factor in reading/comprehensibility. As the number of separate facts and ideas in a reading passage of a given length increases, the likelihood that the learning (comprehension) will be of the "massed practice" type increases. As the number of facts and ideas in a passage of a given length decreases, the learning will be of the "distributed practice" type. Hence, the more the reading selection requires the memory for semantic units factor the greater the probability that the learning will be of the massed rather than of the distributed type.

Distributed and massed practice principles can also be employed to support the contention that increased convergent production of semantic systems will decrease comprehensibility. With regard to the convergent factor, the more parts to the system which the reader has to integrate in a passage of a given length, the more the learning will be "massed." Essentially, the two aforementioned factors can be reduced to one; i.e., the greater the amount of information that the reader must process in a reading selection of a given length, then the greater the probability that "massed practice" will ensue with the consequent decrease in reading comprehension.

Whole vs. Part Learning. When comparing whole learning to part learning, several factors must be taken into consideration:

1. part learning is better when motivation is low, because results appear sooner

2. whole learning is superior with greater continuity and meaningfulness in materials to be learned

3. as the amount of material to be learned increases, the part method increases in usefulness

4. as the intelligence of the readers increases, the whole method increases in superiority

5. after practice with both methods, the whole method gradually assumes superiority (Hovland, 1951)

Given the aforementioned principles, it would seem that whole learning is superior to part learning, except when the material to be learned is quite difficult, lengthy, or lacking in meaningfulness. On this basis, then, if the level of the selected structure-of-intellect factors incorporated into reading material is increased, then the reader is forced to use the less effective part learning method. When confronted with "meaningful" or difficult text, the reader is apt to use the part method to learn (comprehend) the passage. As we have just shown, use of the part method generally results in slower learning.

Serial Learning. Serial learning phenomena, as described by Hovland (1951), are dependent upon the number of isolated facts that the reader is required to learn. In addition, items at the beginning and end of a series are learned more quickly than those at the middle of the series. Seemingly, serial learning phenomena are most relevant to memory for semantic units, convergent production of semantic units, and convergent production of semantic systems. As the number of facts or ideas which must be remembered increases, the longer it will take to learn (comprehend) a given reading passage. Similarly, as the number of parts in a system which the reader is required to assimilate increases, the longer the learning (comprehension) time. In summation, then, serial learning phenomena offer a singularly important insight into the reason why comprehension will be increased if two of the structure-of-intellect operations are considered in the writing of prose materials.

Multiple Discrimination Learning. Multiple discrimination learning principles can be applied to explain why the memory for semantic units factor should influence the comprehensibility of textual materials. According to Hovland (1951) and Gagné (1965), multiple discrimination learning is facilitated by material which is meaningful, distinct, and differentiated. If the material contains many ideas and facts which are relatively undifferentiated (memory for semantic units), then the material will be harder to comprehend than material which contains fewer replicated facts and ideas.

Generally, the more meaningful the material, the easier it is to learn. For example, if one compares time to learn nonsense syllables with the time it takes to learn related words, he will find the latter to be learned more quickly. The meaningful material is more differentiated and the nonmeaningful material is less differentiated and amorphous (Hovland, 1951). Text which is loaded on the evaluation of symbolic units suffers in the same way as nonsense syllables. Symbolic units are less familiar and more abstract representations of ideas and concepts. Since symbolic units are often extreme abstractions which have little meaning in and of themselves, they serve to interfere with learning.

Similarly, as the number of required class members in a text requiring divergent production of semantic units increases, the less the explanatory value or comprehensibility of the text. When a text requires the convergent production of semantic systems and the convergent production of semantic implications, the text is essentially incomplete. When performing convergent production of semantic systems, the reader is required to extrapolate from the material. When performing convergent production of semantic implications, the reader is required to draw inferences or deductions not provided in the text. The incompleteness described in the two foregoing instances is an indication of a relative lack of meaningfulness. This lack of meaningfulness, then, should produce a decrease in the comprehension of textual material.

Stimulus Generalization. The more similar a situation is to that of the original learning situation, the higher the likelihood that it will evoke the same learned response or behavior (Hilgard, 1956). Similarly, in reading, the more the comprehension questions are similar in their stimulus properties to the reading passage, the better the individual is likely to show comprehension of what he has read. This problem can contaminate the measurement of comprehension. The individual may comprehend the reading passage. But, because its stimulus properties are different from the comprehension questions, there may be a lack of stimulus generalization and an apparent lack of learning. This explanation is essentially the same as that offered by Underwood (1949) to describe contextual factors in learning and retention. That is, if one learns under a certain context and is tested for retention under another context, the amount retained will be a function of the degree to which the learning-retention context was changed. Underwood cites an experiment in learning paired nonsense lists, in which only the background color of the paper was changed during the retention test. During the retention test, considerable decrement was evidenced (28 per cent) and relearning took almost three times the number of trials as when the context was unchanged. This stimulus generalization (contextual) aspect of learning can be used to explain why reading materials loaded on memory for figural and, indeed, on all the remaining structure-of-intellect factors, will yield inferior comprehension.

Contiguity and Word Chaining. One of the canons of classical behavior theory is that contiguity is necessary for learning to occur (Gagné, 1965). Contiguity is reflected in the cognition of semantic relations and convergent production of semantic systems. Reading materials which are loaded on these factors fail to incorporate contiguity. In order for the reader to comprehend properly the meaning, he must mentally place the verbal material in proper juxtaposition. He must impose contiguity or systematization on material which is not contiguous or systematic. Since the ability to impose contiguity is dependent on prior experience, information, and logic, we would expect that comprehending reading materials loaded in the cognition of semantic relations and convergent production of semantic systems would be more difficult.

With regard to word chains, most of us have a repertoire of chains of verbal associations that have been previously learned. That is, word ordering and word chains are predictable. Many words within phrases are dependent and tend to occur together (Miller, 1951). Accordingly, when previously learned chains are disrupted, or are not present in reading passages, then, new learning may be required on the part of the reader. Again, as with contiguity, the presence of cognition of semantic relations factor in reading material may disrupt previously learned chains and produce a consequent reduction in reading comprehension.

Motivation. Staats and Staats (1963) suggested that the more intensive the work involved in learning discriminative stimuli in reading, the more aversive the reading will become. That is, the more concentrated effort and work the individual has to put forth, the more unpleasant the reading behavior. Therefore, any response that removes the individual from the aversive learning situation will be reinforced. Although these comments constitute an argument in favor of more gradual introduction of discrimination learning in children, it can also apply to adults who are unexpectedly given a very difficult or unfamiliar (in form) reading passage which requires a considerable amount of work on their part. This motivational factor applies to all of the selected structure-of-intellect comprehensibility factors, inasmuch as loading a text on any or all increases the work effort requirement, and a consequent decrease in motivation is likely to occur.

Additionally, when a more difficult learning (comprehension) task is involved, a motivational problem is apt to become involved. The adult reader may not foresee the probability of obtaining any reinforcement for his extra effort and will therefore work less to learn the materials loaded more heavily on the structure-of-intellect factors.

Gestalt Theory

The gestalt theorists, although primarily concerned with perception, had much to say about learning. One of the basic gestalt laws, the law of prägnanz, states that learning is a restructuring of the field, or a perceptual reorganization in order to form more complete gestalts (Pittenger & Gooding, 1971; Hilgard, 1956; Wertheimer, 1958). The "law of prägnanz" is based on several other "laws." These are discussed in the paragraphs that follow.

Closure. Perhaps the most salient gestalt principle, for our purposes, is that of closure. In perception, a closed figure is one which is bounded (Bobbitt, 1958). In learning, when the whole is not complete, tension arises with an accompanying drive toward completion (closure). This is the gestalt "law of effect" which allows for reinforcement (Hilgard, 1956). Pittenger and Gooding (1971) indicated that unlearned material lacks closure or is ambiguous. The learner normally attempts to reduce the ambiguity which exists. It is not teacher behavior which causes learning, but "Learning is a ... process of organizing perceptions to reduce ambiguity (solve problems)" [p. 97].

When examining the structure-of-intellect based comprehensibility factors, it is apparent that several involve the principle of closure. That is, materials loaded on these factors set up a state of tension or ambiguity which the reader must remove. A textual passage which is loaded heavily on the divergent production of semantic units factor makes the reader strive toward closure. In essence, then, this type of text requires more restructuring of the field on the part of the reader. This task may be quite difficult or impossible for some readers.

Reading materials loaded in the convergent production of semantic systems also fail to produce closure because the reader is required to assimilate the parts (sentences) of the system into a systematic whole.

Finally, reading materials loaded on the convergent production of semantic implications impose a heavy closure involvement. That is, with such materials the reader must cognitively supply the missing information in order to reduce the state of ambiguity.

Similarity. The law of similarity simply indicates that individuals tend to group perceptually similar items together (Wertheimer, 1958). With regard to paired associates learning, the gestalt psychologists demonstrated that similar pairs were more easily learned than dissimilar pairs (Hilgard, 1956).

In one sense, the law of similarity only applies to the memory for semantic units factor. That is, the more differentiated (dissimilar) the units or words in the text, the harder they are to learn or comprehend. In the other sense, the law of similarity can be applied to all of the selected structure-of-intellect factors in the same manner as stimulus generalization. The greater the similarity between the comprehension questions and the textual material, the greater will be the comprehension.

Proximity

The gestalt principle of proximity can be considered equivalent to behavior theory's principle of contiguity. That is, items that are temporally or spatially connected together are considered more meaningful. Perceptually, the learner is compelled to group nonproximate items together into a meaningful whole. Many memory traces based on the law of proximity can be built up in an individual such that he expects certain words to occur together in a specific juxtaposition. This latter concept is equivalent to the word chaining phenomenon of behavior theory. The law of proximity applies primarily to reading materials involving cognition of semantic relations and secondarily to materials involving memory for figural units and evaluation of symbolic units. Text which imposes a heavy cognition of semantic relations load on the reader violates the law of proximity, because many of the words and phrases in the reading passages are not couched in a meaningful manner. The reader is required to restructure the material. This increases the material's difficulty level. The reader must cognitively impose proximity on the material so that it can become meaningful. Consider the following passage:

> The doctor and the psychiatrist entered the hospital room. They proceeded to pull out their syringes, medicines, and Rorschach cards.

The passage violates the law of proximity and the stimulus trace concept based upon the law of proximity. The passage would be more meaningful and require less relational ability if it read:

> The doctor and the psychiatrist entered the hospital room. The doctor proceeded to pull out his syringes and medicines while the psychiatrist pulled out his Rorschach cards.

In the latter passage, the reader is not required to cognitively link the doctor with syringes and the psychiatrist with Rorschach cards. The passage accomplishes this for him. The passage brings the doctor and the psychiatrist in proper juxtaposition to the items they are removing from their bags.

Prose loaded on the symbolic and figural factors can be considered more difficult, in gestalt terms, because of the dearth of stimulus traces for these materials. One is less likely to have built up a series of symbolic or figural traces than to have built up a series of verbal traces. Such material is more difficult because there is a relative lack of past experience with symbols and figures. That is, one cannot reply on familiar word chains and phrases to ease comprehension.

Span of Perception. Miller (1958) has indicated that our ability to process information is limited to seven units, plus or minus two. One exception to this rule is that the span of perception increases with familiar or previously learned material. Since symbolic (and in some cases figural) material is often completely unfamiliar, memory for such units of information will be limited to the typical span of perception. Accordingly, the reader will have to devote more study to symbolic and figural material than to the more familiar conventional prose material in order to reach the same level of comprehension. Given the same amount of time to read both conventional and symbolic material, comprehension of the symbolic (and figural) material will suffer because of the reader's lack of experience and unfamiliarity with the symbolic material.

Miller's thesis may also be applicable to the memory of semantic units. Reading materials will increase in difficulty to the extent that they contain more units of information (facts, ideas, etc.) than the span of perception allows. Of course, given sufficient time to study a passage, familiarity with the contents will increase, thus increasing the span of perception.

<u>Figure Ground and Signal Detection.</u> The concept of figure and ground would seem to be related to several of the selected structure-of-intellect factors. "In relation to ground, the figure is more impressive and more dominant. Everything about the figure is remembered better, and the figure brings forth more associations than the ground" [Ruben, 1958, p. 199]. Certain types of reading presentations tend to confuse figure and ground and make it more difficult to perceive figure. As with many of the other principles here described, familiarity and past experience with various types of material can determine what is perceived as figure and what is perceived as ground. Accordingly, materials requiring the evaluation of symbolic units (since it is unfamiliar) will delay the reader's forming figure and ground concepts. He will have to examine and familiarize himself with the symbolic material before he is able to differentiate figure from ground. Until the figure is differentiated from the ground, no meaning or comprehension of the material can occur. Because the reader has a considerable amount of experience with conventional prose, he is more apt to be able to differentiate figure (grasp the meaning of) from ground (the stimulus constellation) when such conventional prose is used.

In reading material which is loaded on the divergent production of semantic units, there are fewer class members included in the text. This lack of class members is analogous to a lack of figure. Accordingly, we hold that it will be more difficult for the reader to comprehend such material.

The convergent production of semantic systems ability is also related to the figure and ground concept. The reader is presented with a mass of unsystematic information from which he must extract figure (assimilate, integrate, etc.). The reader is required to derive order from such prose before it can be comprehended. Theory of signal detection is based on the observer's ability to detect or differentiate signal when presented with both signal and noise. The "hit" rate is the proportion of time the observer reports signal when signal is present. The "false alarm" rate is the proportion of time the observer reports signal when noise is present. The "miss" rate is the proportion of time the observer reports noise when signal is present. All of the selected readability/comprehensibility factors obfuscate or make it more difficult to detect signal (derive meaning) when signal and noise are present. Undifferentiated and nonmeaningful material would be considered noise by a reader while differentiated meaningful material would be considered signal by the reader. The extent of reading comprehension, then, can be considered the ratio of the signal strength to the noise strength. Conceivably, material which is completely understood would be signal without noise.

Phenomenological Theory

Combs and Syngg (1959) represent the phenomenological approach to learning. Their thesis is that we can change our behavior only as a result of changes in self perception and changes in how we perceive the environment. From this it follows that readers will fail to learn materials which have no relevance or meaning to their present lives. Therefore, the more obstruse the reading material, or the more it is presented in a foreign manner, the less likely it is to be learned. Materials which are irrelevant to the individual will not exist in his field of experience and he will not be aware of them. The work of Ebbinghous which indicated nonsense syllables to be harder to learn and more easily forgotten supports the phenomenological point of view. Certainly, text which is loaded on any of the structure-of-intellect factors here involved can be considered to violate the phenomenological point of view. Each of the factors, when included in reading matter, tends to make that material either more abstract or foreign to the reader in terms of his past experience and self perceptions. Consequently, the reader will be less able to comprehend material to the extent that it does not fit his relevant experiences. Phenomenological psychologists would probably most heavily emphasize the symbolic and figural factors, since these tend to involve the most abstract and foreign reading materials.

Methods and Procedures

Hypotheses and Experimental Design

The working hypothesis for this phase of the research was that loading reading material on an SI factor will tend to make the material less comprehensible or more difficult to read. Accordingly, when the SI factor is not required or included in a reading selection, the material should be relatively easy to understand. The basic research paradigm was to present two equivalent groups with two reading selections each of which contained exactly the same information. The selections for one of the groups were not heavily loaded on an SI factor; the selections for the other group were highly loaded on the SI factor.

In order to determine if there are differences in the readability/comprehensibility of the two types of reading material, a test of comprehension was employed. The test materials were exactly the same across experimental conditions. This procedure is permissible, inasmuch as the same information was presented in both the high SI load and the low SI load materials.

The question answered by this procedure is whether or not varying the SI load imposed by a reading selection accounts for a significant proportion of the variation in reading comprehension test performance. If those individuals who read the material which was highly loaded on the SI factor scored significantly lower on the test than those individuals not required to read the SI loaded material, then the hypothesis is confirmed.

Relevant Structure-of-Intellect (SI) Factors

The eight selected SI factors which were considered relevant to the readability/comprehensibility of written material are reviewed in Table 2-1.

Table 2-1

Eight Selected Guilford SI Factors and Their Associated Acronyms

SI Factor	Acronym
Cognition of Semantic Units	CMU
Cognition of Semantic Relations	CMR
Memory of Figural Units	MFU
Memory of Semantic Units	MMU
Convergent Production of Semantic Implications	NMI
Convergent Production of Semantic Systems	NMS
Divergent Production of Semantic Units	DMU
Evaluation of Symbolic Units	ESU

The reading selections employed and their associated tests are presented and discussed in the subsequent sections of this chapter.

Cognition of Semantic Units (CMU)

According to Guilford (1967), the CMU factor is best measured by tests of vocabulary. Our readability/comprehensibility conjecture relative to this factor was based on vocabulary diversity. The type/token (T/T) ratio was chosen as the index of vocabulary diversity. The T/T ratio is defined as the ratio of the number of different words (types) to a total number of words (tokens).

In order to provide a reading selection with a low CMU involvement, a section from a children's encyclopedia written at the fourth grade level was selected. A sample portion of this reading selection is presented in Figure 2-2.

Light

Have you ever heard someone say, "Turn the light on--I can't see a thing?" Or: "We'll have to wait until the sun rises before we can see?" .

Without light we would be lost. A long time ago people depended upon the light of the sun to do their work. They would begin to work when the sun rose and would stop when the sun set.

Then people discovered fire and found that it could light rooms at night. You know the famous story about how Abe Lincoln used to read a great deal in front of his fire-place just to get the light from the fire. Of course, many people used candles, if they could afford them. Later a fuel--kerosene--was used in special lamps.

Still later, gas--illuminating gas--gave us light in our homes and even in our streets.

Figure 2-2. A sample from the Cognition of Semantic Units (CMU) reading selection involving a low CMU load on the reader.

The approach used to increase the T/T ratio involved systematic changes in the wording of the selection. Whenever possible a synonym was used to replace some of the words that were used more than once. Occasionally, entire phrases were changed in-order to introduce variability. In all cases, care was exercised so that the meaningfulness of the selection remained the same. A corresponding high CMU load sample portion from this reading selection is shown in Figure 2-3.

Light

Did you ever hear someone say, "Turn the light on--I can't see a thing?" Or: "We'll have to wait until the sun rises before we can visualize?"

Without illumination people would be lost. A long time ago the populace depended upon electromagnetic radiation from a star to perform their tasks. Individuals would begin their work when sol rose and these persons would discontinue when sol set.

Then fire was discovered and the natives learned that it could illuminate rooms at night. You know the famous story about how Abe Lincoln used to read a great deal in front of his fireplace in order to obtain the illumination given off by the flames. Of course, many persons lit candles, if they could afford these objects. Later a fuel--kerosene--was burned in special lamps.

Subsequently, a gaseous compound--illuminating gas--gave us light within our abodes and even upon our streets.

Figure 2-3. Sample from the Cognition of Semantic Units (CMU) reading selection involving a high CMU load on the reader.

The T/T ratio for the selection involving little in the way of the SI factor was .504. The T/T ratio for the selection loaded heavily on the SI factor was .611. This sort of difference in T/T ratios is considered sizeable by conventional standards. Some sample items from the test of the CMU factor are presented in Figure 2-4.

Directions

Please fill in the blank spaces with the most correct or appropriate word(s) or phrases.

1. ________________________ was the first source of light used by people to do their work.

2. ________________________ used to read by the fireplace

3. People discovered ____________________ which was used to light rooms at night.

4. People who could afford them used __________________ as a source of light.

5. ________________________ was the first fuel in special lamps.

Figure 2-4. Sample test items from the Cognition of Semantic Units Test.

Cognition of Semantic Relations (CMR)

Cognition of semantic relations is the ability to recognize the relation between two items or words. Guilford (1967) used analogy and word linkage tests to measure this factor. Accordingly, the reading selections which involve the CMR factor require the reader to form word linkages and tax his ability to form correct relations Figures 2-5 and 2-6 present portions of the reading selections used to vary the CMR factor. The sample selection in Figure 2-5 provides the word linkages, whereas the sample selection in Figure 2-6 requires the reader to form the correct linkage or relation. For example, the first linkage in Figure 2-5 involves delivery of maps to a sergeant and doughnuts to the cooks in the mess hall. In Figure 2-6, the correct mops-sergeant and doughnuts-cooks linkage is not provided for the reader. Accordingly, this places a heavier mental load on the reader.

Truck Driving

When George Farguahar joined the Air Force, he did not realize that he would get to see almost every aspect of Air Force base operations. You see, Airman Farguahar was assigned to the base motor pool as a deliveryman. A typical day in the life of Airman Farguahar will be described in order to show how a deliveryman can learn about Air Force base operations.

At 0500 hours Airman Farguahar arrived at the motor pool. He then drove to the base warehouse to begin loading his truck. By 0530 the truck was loaded for the morning deliveries.

The first delivery in the morning was to the 43rd Squadron Mess Hall. A delivery had to be made to the cooks and to the sergeant in charge of the clean up detail at this mess hall. Airman Farguahar delivered mops to the sergeant and doughnuts to the cooks. Next, Airman Farguahar went to the base carpenters and machinist shop in order to deliver nails to the carpenters and wrenches to the machinists. The next delivery was to the automotive repair shop to which torque wrenches were delivered to the auto mechanics and fiberglass putty was delivered to the body repairmen. At 0900 hours Airman Farguahar took a much needed coffee break.

Figure 2-5. Sample from the Cognition of Semantic Relations (CMR) reading selection involving a low CMR load on the reader.

Truck Driving

When George Farguahar joined the Air Force he did not realize that he would get to see almost every aspect of Air Force base operations. You see, Airman Farguahar was assigned to the base motor pool as a deliveryman. A typical day in the life of Airman Farguanar will be described in order to show how a deliveryman can learn about Air Force base operations.

At 0500 hours Airman Farguahar arrived at the motor pool. He then drove to the base warehouse to begin loading his truck. By 0530 the truck was loaded for the morning deliveries.

The first delivery in the morning was to the 43rd Squadron Mess Hall. A delivery had to be made to the cooks and to the sergeant in charge of the clean up detail. Airman Farguahar delivered their mops and doughnuts. By 0800 hours he was already on his way to the base carpentry and machinist shop in order to deliver to them their supplies of nails and wrenches. The next delivery was to the automotive repair shop at which a delivery had to be made to the auto mechanics and body repairmen. At 0900 hours Airman Farguahar delivered their special torque wrenches and fiberglass putty. At 0945 Airman Farguahar took a much needed coffee break.

Figure 2-6. Sample from the Cognition of Semantic Relations (CMR) reading selection involving a high CMR load on the reader.

The metric for the CMR factor involved a tabulation of the number of linkages that the reader is required to form per 100 words. In the low CMR selection, the value of this metric was 0.00/100 words while in the more difficult selection the value was 1.98/100 words.

Figure 2-7 presents a sample of the test questions for this factor.

Truck Driving

Directions

Please fill in the blank spaces with the most correct or appropriate word(s).

1. Airman Farguahar delivered ________________ to the sergeant in charge of the cleanup detail.

2. ______________________ to the cooks of the mess hall.

3. Airman Farguahar delivered ________________ to the carpenters and

4. ______________________________ to the machinists.

5. _________________________ was delivered to the body repairmen at the automotive shop.

6. __________________________ was delivered to the auto mechanics at the automotive shop.

Figure 2-7. Sample test items from the Cognition of Semantic Relations test.

Memory of Figural Units (MFU)

Map reading was considered the most relevant type of material for measurement of the MFU factor (Guilford & Hoepfner, 1971). It was hypothesized that as more information (in the form of labelled locations and items) is presented on a map, the more difficult it will be to remember specific aspects of the map.

Figure 2-8 presents the map employed to present only the required amount of relevant information to the reader. Figure 2-9, on the other hand, presents the map which contains excess information. Accordingly, the map presented in Figure 2-9 places a heavier MFU load on the reader than the map presented as Figure 2-8.

The metric employed to measure this factor was based on a tabulation of the number of labelled locations. The map in Figure 2-8 contains 43 labelled locations whereas the map presented as Figure 2-9 contains 92 labelled locations. Sample items for the MFU factor are shown in Figure 2-10.

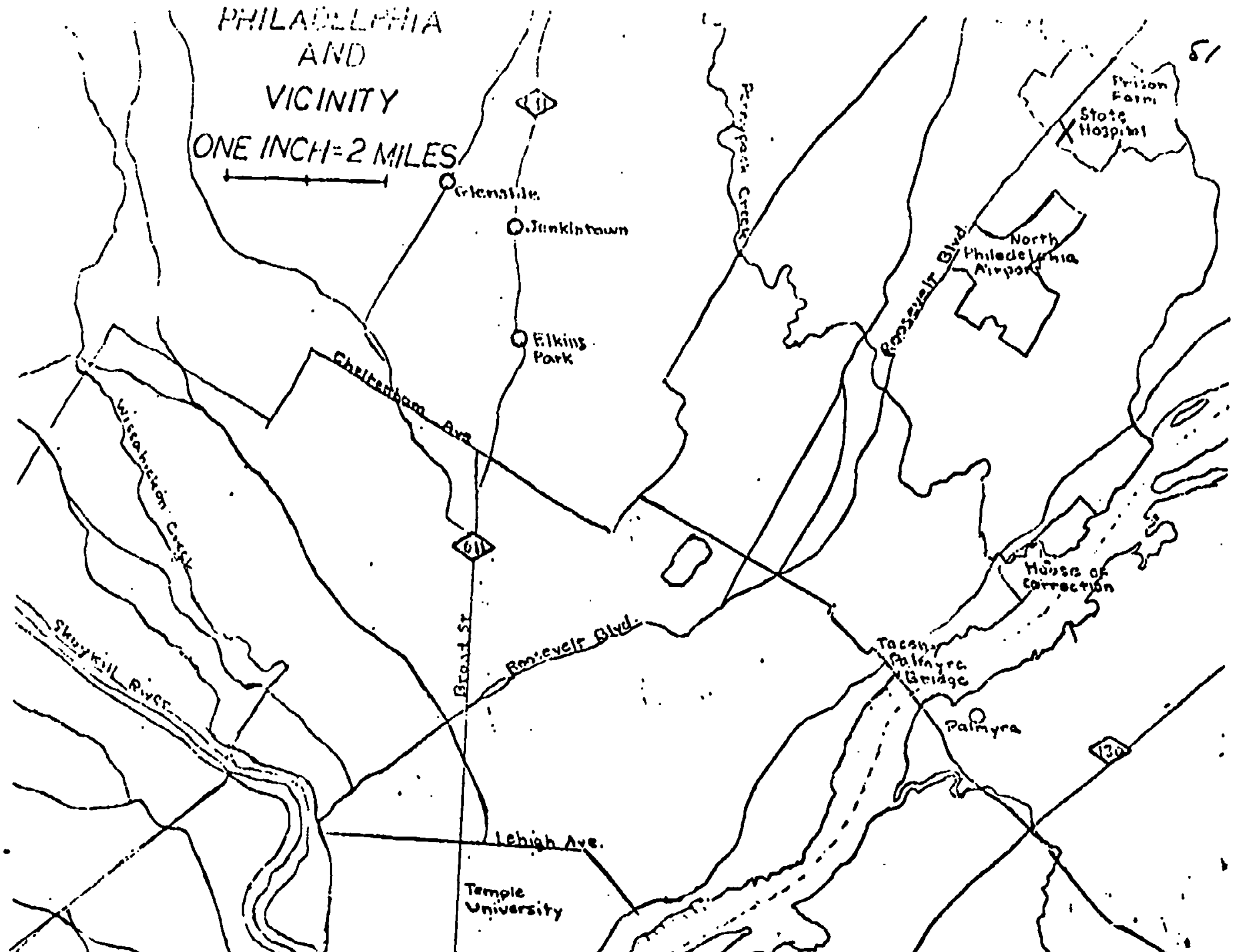

Figure 2-8. Memory of figural units (MFU) reading selection involving a low MFU load on the reader.

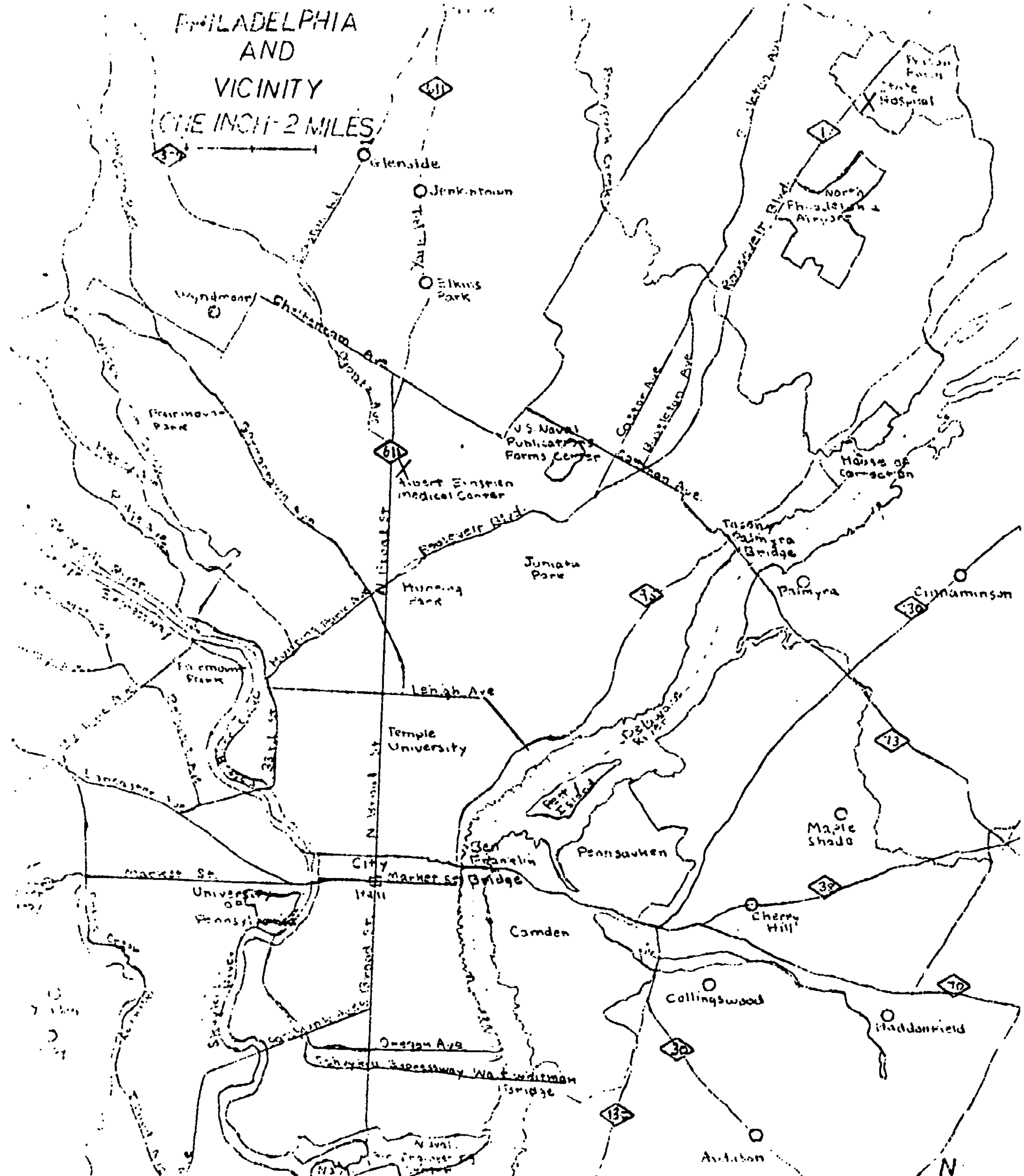

Figure 2-9. Memory of figural units (MFU) reading selection involving a high MFU load on the reader.

Directions

Circle the T in front of the statement if it is true. Circle the F in the front of the statement if it is false.

1. T F The state hospital is located in the extreme northeast section of Philadelphia.

2. T F Market Street is the main north and south thru street in Philadelphia.

3. T F One can cross from Philadelphia to New Jersey by using either of two bridges.

4. T F Temple University is located in the approximate geographical center of Philadelphia.

5. T F The Schuykill is the largest river that cuts through Philadelphia.

Figure 2-10. Sample test items from the Memory of Figural Units test.

Memory of Semantic Units (MMU)

Typically, tests of memory for ideas are used to measure the MMU factor (Guilford, 1967). The reading selection which was loaded with a relatively greater amount of MMU had fewer replicated ideas and facts than the selection less highly loaded. Figures 2-11 and 2-12, respectively, present samples of the MMU reading selections involving high and low MMU loads. The metric employed in this instance was the number of replicated facts per 100 words--4.25 and .27 for the selections presented in Figures 2-11 and 2-12, respectively. In this case, a higher value indicates easier material. Since it takes more words to write a paragraph with replicated facts, filler material was added to the selection involving the greater MMU load. This procedure equated the length of the easy and the difficult materials.

Electronics Ground Safety

Medical records prove that electrical currents great enough to cause actual burning kill less often than do currents of much lower magnitude. In other words, currents of lower magnitude kill more often than larger currents which are capable of burning. Electricity kills by overriding the control that the nervous system exercises from controlling bodily functions. The human body has sometimes been compared to an automatic factory. Muscles are the motors of this human automatic factory. Masterminding the operation of these muscle motors of the human body is that fabulously complicated calculator--the brain. The brain, then, controls the operation of the muscle motors of the body. This message center sends instructions to the controlled parts of the body through an intricate electrochemical network we know as the nervous system.

If overriden by an outside current, the electrical impulses of the nervous system lose control of body functions. External electric currents, then can result in the nervous system losing control of the body. Particularly dangerous are currents that enter the heart and respiratory centers. Heart and respiratory centers are particularly vulnerable to electrical currents, because they are vital to body functions. Thus, a key factor in death by electrical shock is the path of the undesired current within the human body, as well as its magnitude.

Figure 2-11. Sample from the Memory for Semantic Units (MMU) reading selection involving a low MMU load on the reader.

Electronics Ground Safety

Medical records prove that electrical currents great enough to cause actual burning kill less often than do currents of much lower magnitude. Electricity kills by overriding the control that the nervous system exercises over the body. The human body has sometimes been compared to an automatic factory. Muscles are its motors. Masterminding the operation of these motors is that fabulously complicated calculator--the brain. This message center sends instructions to the controlled parts of the body through an intricate electrochemical network we know as the nervous system.

If overriden by an outside current, the electrical impulses of the nervous system lose control of the body. Particularly dangerous are currents that enter the heart and respiratory centers. Thus, a key factor in death by electrical shock is the path of the undesired current within the human body, as well as its magnitude.

(Filler material added to equate for length.)

Figure 2-12. Sample from the Memory for Semantic Units (MMU) reading selection involving a high MMU load on the reader.

Electronics Ground Safety

Directions

Please fill in the blank spaces with the most correct or appropriate word(s).

1. Currents that can burn kill________________ smaller currents.

2. Electricity kills by overriding the_________ that the

3. ______________________________exercises over the body.

4. The human body has sometimes been compared to an__________________________________.

5. The___________________________masterminds the operations of the muscles.

Figure 2-13. Sample test items from the Memory of Semantic Units test.

Convergent Production of Semantic Implications (NMI)

Guilford and Hoepfner (1971) used syllogisms as a measure of the NMI factor. Reading material incorporating a high syllogistic reasoning requirement might look like that shown in Figure 2-15. Figure 2-14 presents a sample of the same material with the required syllogisms completed for the reader.

Airman Work

Each airman in Squadron A is required to perform K.P. duty at least once in any calendar month. The only exceptions to this regulation are sickness, emergency leave, or guard duty. With regard to each of these exceptions, the designated airman is required to make up his missed K.P. duty so that another airman will not have to take K.P. twice in his stead. Any airman who is assigned K.P. duty twice in any month is not required to take it the following month.

Airman Smith has not taken K.P. duty as of April 29th of this year. If April has 30 days, and Airman Smith is not on guard duty or is not sick, then we know that Airman Smith will take K.P. duty on April 30th or that Airman Smith is on emergency leave. We also know if Airman Smith is on emergency leave, that he will take K.P. duty twice during the month of May.

Airman Johnson was assigned K.P. duty twice during the month of April. Airman Johnson had K.P. duty on April 10th and on April 30th. We know, then, that Airman Johnson was taking K.P. for someone who was either sick, on emergency leave, or on guard duty. We also know that Airman Johnson will not have to take K.P. during the month of May. We do not know, though, that Airman Johnson is taking K.P. duty for Airman Smith, since Airman Johnson may have been taking K.P. for someone other than Airman Smith.

Airman Lockhart was on emergency leave from April 1st to April 20th. On April 21st Airman Lockhart was on guard duty and on April 25th Airman Lockhart reported in sick to the infirmary. We can not assume from the above information that Airman Lockhart missed his K.P. duty for the month of April, inasmuch as there were several days remaining in April in which Airman Lockhart could have served his K.P. duty.

Figure 2-14. Sample from the Convergent Production of Semantic Implications (NMI) reading selection involving a low NMI load on the reader.

Airman Work

Each airman in Squadron A is required to perform K.P. duty at least once in any calendar month. The only exceptions to this regulation are sickness, emergency leave, or guard duty. With regard to each of these exceptions, the designated airman is required to makeup his missed K.P. duty so that another airman will not have to take K.P. twice in his stead. Any airman who is assigned to K.P. duty twice in any month is not required to take it the following month.

Airman Smith has not taken K.P. duty as of April 29th of this year. The month of April has 30 days.

Airman Johnson was assigned K.P. duty twice during the month of April. Airman Johnson had K.P. duty on April 10th and April 30th.

Airman Lockhart was on emergency leave from April 1st to April 25th. Airman Lockhart reported in sick to the infirmary.

Figure 2-15. Sample from the Convergent Production of Semantic Implications (NMI) reading selection involving a low NMI load on the reader.

As in the MMU factor, the high level NMI factor reading selection required the addition of filler material, inasmuch as the harder material with the incomplete syllogisms consisted of fewer words. The NMI metric was the number of syllogisms required of the reader per 100 words. The reading selection requiring little of the NMI ability (Figure 2-14) obtained a metric score of .00, while the reading selection requiring relatively more of the NMI (Figure 2-15), obtained a metric score of 1.65.

Figure 2-16 presents sample test items from the NMI test.

Directions

Please circle the T next to the statement if it is true. Circle the F if the statement is false. Circle the U next to the statement if it can not be determined whether the statement is true or false based upon the information given in the text.

1. T U F If Airman Smith was on emergency leave on April 30th would he be required to take K.P. duty twice during the month of May?

2. T U F Airman Smith will take K.P. duty on April 30th.

3. T U F Airman Johnson was taking K.P. duty for Airman Smith.

4. T U F Airman Johnson was taking K.P. duty for someone else other than Airman Smith.

5. T U F Airman Lockhart missed his K.P. duty for the month of April.

6. T U F Airman Smith will take K.P. on April 30th unless he is on guard duty, has emergency leave, or is sick.

7. T U F Airman Johnson was taking K.P. for either Airman Smith or for someone else other than Airman Smith.

8. T U F Airman Lockhart served K.P. duty between April 22nd and April 24th.

Figure 2-16. Sample test items from the Convergent Production of Semantic Implications Test.

Convergent Production of Semantic Systems (NMS)

In the measurement of the NMS factor Guilford (1967) used tests of ordering or organizing ability. In the present context, one well ordered and highly structured selection was constructed (low NMS load). Various tabular presentations, mnemonic devices, and visual aids were incorporated into this highly structured selection with the view that they would aid the subject in organizing and learning the selection. Conversely, another passage (high NMS load) was constructed which contained the same relevant information, but which lacked the various ordering devices present in the selection described first. Figure 2-17 presents a sample of the well ordered material requiring little of the NMS ability, while Figure 2-18 presents a sample of the less well ordered prose.

Artificial Respiration

MOUTH-TO-MOUTH METHOD (EXHALED-AIR METHOD). It has been proven that the mouth-to-mouth method of artificial respiration is the most effective method.

It is simpler to use and saves more lives. Don't waste time trying old methods, or worrying about getting infected. The possibility of infection is remote. YOU HAVE A LIFE TO SAVE.

In this method, you breathe air into the victim's lungs with your own mouth. Since you consume only part of the oxygen out of the air which you inhale, the air you breathe into the victim's lungs contains enough oxygen to revive him.

You'll find that you need to breathe slightly deeper and faster than usual in order to get enough air for yourself, but don't worry about this point. Under certain conditions, which will be explained later, the mouth-to-mouth method of artificial respiration cannot be used. The step-by-step procedure for administering mouth-to-mouth artificial respiration follows:

STEP 1. TURN THE VICTIM ON HIS BACK.
STEP 2. CLEAN THE MOUTH, NOSE, AND THROAT.
STEP 3. PLACE THE VICTIM'S HEAD IN THE "SWORD-SWALLOWING POSITION."
STEP 4. HOLD THE LOWER JAW UP.
STEP 5. CLOSE THE VICTIM'S NOSE.
STEP 6. BLOW AIR INTO THE VICTIM'S LUNGS.
STEP 7. LET THE AIR OUT OF THE VICTIM'S LUNGS.

Repeat Steps 6 and 7 at a rate of 12 to 20 times per minute. Continue rhythmically without interruption until the victim starts breathing or is pronounced dead. A smooth rythym is desirable but split-second timing is not essential.

An easy way to remember this sequence is to divide the steps according to the key words, and then remember the key words in pairs or triplets. For example:

Key Word	STEP
Turn	1
Clean	2
Place	3
Hold	4
Close	5
Blow	6
Out	7

The table shows one triplet of key words and two pairs of key words which work as good aids in remembering this seven step procedure.

Figure 2-17. Sample from the Convergent Production of Semantic Systems (NMS) reading selection involving a low NMS load on the reader.

Artificial Respiration*

MOUTH-TO-MOUTH METHOD (EXHALED-AIR METHOD). It has been proven that the mouth-to-mouth method of artificial respiration is the most effective method.

It is simpler to use and saves more lives. Do not waste time trying old methods, or worrying about getting infected. The possibility of infection is remote. YOU HAVE A LIFE TO SAVE.

In this method, you breathe air into the victim's lungs with your own mouth. Since you consume only part of the oxygen out of the air which you inhale, the air you breathe into the victim's lungs contains enough oxygen to revive him.

You'll find that you need to breathe slightly deeper and faster than usual in order to get enough air for yourself, but don't worry about this point. Under certain conditions, which will be explained later, the mouth-to-mouth method of artificial respiration can not be used. The step-by-step procedure for administering mouth-to-mouth artificial respiration follows: turn the victim on his back; clean the mouth, nose and throat; place the victim's head in the "sword-swallowing position"; hold the lower jaw up; close the victim's nose; blow air into the victim's lungs; let air out of the victim's lungs.

Repeat the last two steps at the rate of 12 to 20 times per minute. Continue rythmically without interruption until the victim starts breathing or is pronounced dead. A smooth rythym is desirable, but split-second timing is not essential.

Figure 2-18. Sample from the Convergent Production of Semantic Systems (NMS) reading selection involving a high NMS load on the reader.

---

*From AF Student text 3AQR30030

The metric for these selections was the number of words, items, or phrases that required ordering per 100 words. For the total reading selection which required a minimal amount of the Guilford ability, the metric was zero. On the other hand, the metric for the selection requiring a relatively greater amount of the Guilford Ability was 1.79. Some sample test items from the NMS test are shown in Figure 2-19.

Artificial Respiration

Directions

In most of the following questions you will be asked to construct lists of various items. Within these lists, the correct answer as well as the sequence of answers is considered important. Accordingly, you will receive two points credit for each correct answer plus an additional point for correct sequential placement.

1. List the steps for administering mouth-to-mouth artificial respiration. (Make certain that you number each step).

   ____
   ____
   ____
   ____
   ____
   ____

2. Which steps in mouth-to-mouth resuscitation are repeated? (Write out the steps).

   ____
   ____

Figure 2-19. Sample test items from the Convergent Production of Semantic Systems test.

Divergent Production of Semantic Units (DMU)

Guilford (1967) indicated that DMU involves the ability to enumerate class members given certain class properties. Accordingly, the reading selection which was low on DMU did not require the reader to enumerate class members. A portion of this reading selection is presented in Figure 2-20. The other, more difficult reading selection, required the reader to enumerate the class members. A sample from this latter selection is presented in Figure 2-21.

The metric for measuring the level of the DMU requirement imposed by the text was the number of divergent productions required of the reader per 100 words of text. For the material not loaded on this factor, the metric value was .00, whereas the metric for the material requiring the factor was 1.40.

Again, since the more difficult material, requiring enumeration of class members (divergent production), is somewhat shorter in length some filler material was added in order to control for this variable.

Sample test items used to measure comprehension of the DMU reading selections are shown in Figure 2-22.

Central Tendency

The mean is the arithmetic average of a group of scores. It is obtained by adding together all the scores in your sample and dividing by the number of persons (N) in the sample. The notation for the mean of all raw scores is usually $\overline{x}$. A small x is usually used as the notation for a single raw score. The Greek letter (Σ) indicates the arithmetic operation of addition. Your formula for computing the mean, then, reduces to:

$$\overline{x} = \frac{\Sigma x}{N}$$

The value for $\overline{x}$ increases with increases in the value of Σx. On the other hand, the value for $\overline{x}$ decreases when N gets larger.

Another measure for central tendency is the median. The median is the midpoint or middle score of a set of scores when the scores are arranged from lowest to highest. When there is an even number of scores, the median is the average of the two middle scores. The following example illustrates this point.

Example

4
5
6
7
← *Median= 7.5*
8
10
12
13

The mode is the most common or frequent score in a set of scores. In most cases you will want to find the computed mode. You determine the computed mode by doubling the mean and subtracting this value from three times the median, or:

Computed mode= (3 x median) - (2 x mean)

As the median increases in size the computed mode also increases in size, or if the mean decreases in size the computed mode decreases in size.

Figure 2-20. Sample from the Divergent Production of Semantic Units (DMU) reading selection involving a low DMU load on the reader.

Central Tendency

The mean is the arithmetic average of a group of scores. It is obtained by adding together all the scores in your sample and dividing by the number of persons (N) in the sample. The notation for the mean of all raw scores is usually $\overline{X}$. A small x is usually used as the notation for a single raw score. The Greek letter (Σ) indicates the arithmetic operation of addition. Your formula for computing the mean, then reduces to:

$$\overline{X} = \frac{\Sigma x}{N}$$

The mean is considered by statisticians to be the most sophisticated measure of central tendency. Another measure of central tendency is the median. The median is the midpoint or middle score of a set of scores when the scores are arranged from lowest to highest. The mode on the the other hand is the most common or frequent score in a set of scores. In most cases you will want to find the computed mode. You determine the computed mode by doubling the mean and subtracting this value from three times the median. You will have little occasion to use these latter two measures of central tendency, since most statistics require the use of means rather than medians or modes. The only instance when a median or mode is preferred over a mean is when the score distribution is highly skewed or distorted.

(Filler material added to equate for length.)

Figure 2-21. Sample from the Divergent Production of Semantic Units (DMU) involving a high DMU load on the reader.

Directions

Please fill in the blank spaces with the most correct or appropriate word(s) or phrases.

1. As the value for Σx increases the value for $\bar{x}$________.

2. As N gets larger the value for $\bar{x}$ will________________.

3. When there is an even number of scores, the median is the____________________of the two middle scores.

4. If the mean decreases in size the computed mode will ____________________________in size.

5. As the median increases in size the computed mode ________________________in size.

Figure 2-22. Sample test items from the Divergent Production of Semantic Units test.

Evaluation of Symbolic Units (ESU)

Guilford (1967) used abbreviations tests to measure ESU ability. Accordingly, one of the reading selections used in the present work was heavily loaded in abbreviations, whereas the other selection did not contain such abbreviations. This Guilford factor is particularly relevant in the present context because of the prevalent use of acronyms and abbreviations in military writing. A part of the reading selection which contained no abbreviations is presented as Figure 2-23 while a part of the reading selection which contained abbreviations is presented as Figure 2-24.

The metric for the ESU factor was the number of abbreviated words per 100 words of text. For example, the acronym AFHRL would count as five abbreviated words. The metric, so calculated, for the abbreviated material was 11.77 per 100 words and the metric for the traditional prose was .00 per 100 words.

Airman Smith

Airman Smith is currently in the personnel awaiting training status pool at Keesler Air Force Base. For his first two days among the personnel awaiting training assignment at Keesler Air Force Base, Airman Smith was assigned to Charge of Quarters and Kitchen Police duties. Since no Commander's Week personnel were available, Airman Smith, like most of the other personnel awaiting training assignments at Keesler Air Foce Base, was disgruntled at having to perform Charge of Quarters and Kitchen Police duties. The lack of an adequate number of Commander's Week personnel has resulted in a steady assignment of personnel awaiting training status to these duties.

A week after his Kitchen Police and Charge of Quarters duties, Airman Smith learned that the Commandant of Troops instructed all units at Keesler Air Force Base to reevaluate Charge of Quarters and Kitchen Police requirements. This was done by the Commandant of Troops in order to ensure the integrity of the personnel awaiting training status program. These revised requirements were to be submitted to the Commandant of Troops as soon as possible.

Figure 2-23. Sample from the Evaluation of Symbolic Units (ESU) reading selection involving a low ESU load on the reader.

Airman Smith

Airman (Amn.) Smith is currently in the personnel awaiting training status (PATS) pool at the Keesler Air Force Base (KAFB). For his first two days among the PATS at KAFB, Amn. Smith was assigned to CQ and KP duties, since no Commander's Week (CW) personnel were available. Amn. Smith, like most of the other PATS at KAFB, was disgruntled at having to perform CQ and KP duties. The lack of an adequate number of CW personnel has resulted in a steady assignment of PATS personnel to these duties.

A week after his KP and CQ duties, Amn Smith learned that the Commandant of Troops (COT) instructed all units at KAFB to reevaluate CQ and KP requirements. This was done by the COT in order to ensure the integrity of the PATS program. These revised requirements were to be submitted to the COT as soon as possible (ASAP).

Figure 2-24. Sample from the Evaluation of Symbolic Units (ESU) reading selection involving a high ESU load on the reader.

Figure 2-25 presents a sample of the test questions relating to the content of the ESU reading selections.

Directions

Please fill in the blank spaces with the most appropriate word(s) or phrases. DO NOT USE ABBREVIATIONS!

1. Airman Smith was in the ____________ pool at
2. ____________ Air Force Base.
3. Airman Smith was upset about being assigned to ____________
4. and ____________ duties. This was due to
5. the fact that there were insufficient ____________ personnel available.
6. The ____________ instructed all units at this Air Force Base to reevaluate their requirements.

Figure 2-25. Sample test items from the Evaluation of Symbolic Units test.

Experimental Setting and Time Limits

The reading selections and associated tests were administered in two large testing rooms at the AFHRL Personnel Research Division at Lackland AFB, Texas. Approximately 65-67 persons were tested in each room. One random half of the subjects in each testing room received highly loaded materials and the other random half received the materials which were less loaded on the SI variables. The DMU, MMU, ESU, and CMR selections and tests were administered in one room while the CMU, NMI, CFU, and NMS selections and tests were administered in the other room. One proctor from Applied Psychological Services and two proctors from the Air Force administered the reading selections and tests in each room.

Table 2-2 presents the time limits allowed for reading each selection and completing each of the associated tests. The time limits were found to be adequate. All persons were able to finish all the reading selections and the associated tests.

The personnel involved were a random selection of new recruits just entering basic training at Lackland Air Force Base.

Table 2-2

Time Limits for Each Structure-of-Intellect Reading Selection and Associated Test Administered to Air Force Recruits at Lackland Air Force Base, Texas

SI Factor	Reading Time Limit	Test Time Limit
CMU	20	20
CMR	15	15
MFU	5	10
MMU	20	18
NMI	15	20
NMS	20	25
DMU	20	25
ESU	15	15

Results and Discussion

Introduction

The data analysis rested on several statistical techniques. One aspect of the analytic plan was correlational in nature. The basic hypothesis involved was that a high or low level in reading material of a specific Guilford factor accounts for a significant proportion of comprehensibility test variance. That is, a high or a low level of a specific Guilford factor in text is statistically associated with the comprehension test score for the text. Point-biserial correlation coefficients between the factor requirement (high or low condition) and comprehension test scores wére calculated.

In addition to the point-biserial correlations, t-tests of significance of the difference between the comprehensibility in test scores of the two levels of textual material for each SI factor were calculated.

Means and Standard Deviations

Table 2-3 displays the means and standard deviations of the test scores for both the factor high and the factor low conditions.

Table 2-3

BEST COPY AVAILABLE

Mean, Standard Deviation, and Number of Subjects Completing Eight Tests Based upon High and Low Readability/Comprehensibility Conditions

Factor High				Factor Low			
Factor	N	Mean	σ	Factor	N	Mean	σ
CMU	33	12.18	4.80	CMU	32	14.59	4.00
CMR	33	11.18	4.57	CMR	34	16.62	3.49
MFU*	29	15.55	2.75	MFU	33	18.70	1.99
MMU	33	9.70	5.22	MMU	34	13.12	3.92
NMI	33	10.48	3.02	NMI	33	12.82	2.77
NMS	33	31.70	11.68	NMS	33	39.33	10.23
DMU	33	7.27	4.42	DMU	34	10.71	3.69
ESU	33	20.15	10.26	ESU	34	25.97	7.79

**Although 33 subjects were tested in the MFU factor high condition, four were from the Philadelphia area and were eliminated from the sample.*

Point-Biserial Correlations and t-Tests

Table 2-4 presents the point-biseral correlation coefficients and the associated t-test (statistical significance of the difference between mean value for each of the eight readability/comprehensibility factors).

The point-biseral correlations between readability/comprehensibility factor "high" or "low" and comprehension test score were all exceptionally high. The maximum obtainable point biseral correlation is .80. This reinforces the contention of correlational adequacy.

In regard to the t-tests, one tailed tests were conducted since the directionality of the results was predicted in advance. Each of the hypotheses was confirmed at least at the .025 level of significance. Increasing textual loading on each of the readability/comprehensibility factors to an appreciable degree decreased reading comprehension, as measured.

Table 2-4

Point-Biserial Correlations and t-Test Values for Each of the Eight Readability/Comprehensibility Factors

Factor	t-value	Point-Biserial r	Significance Level-t
CMU	2.17	.29	$p < .025$
CMR	5.39	.56	$p < .002$
MFU	5.03	.54	$p < .001$
MMU	2.97	.35	$p < .005$
NMI	3.24	.39	$p < .001$
NMS	2.78	.33	$p < .005$
DMU	3.41	.39	$p < .001$
ESU	2.56	.31	$p < .001$

The above results clearly indicate that the global and contextual aspects of reading matter, as represented by the selected readability/comprehensibility factors account for a considerable proportion of readability/comprehensibility variance. Accordingly, it seems that the Air Force should consider incorporation of these findings into a readability/comprehensibility program for assessing their textual materials. Only a few simple rules need be followed in order to take advantage of these findings.

It might be argued that the aforementioned highly significant results could be accounted for by differences in reading grade level of the recruits involved in the factor high and the factor low conditions. To investigate this possibility, the General (G) score of the Airman Qualifying Exam (AQE) can be converted into a Reading Grade Level (RGL) in accordance with a formula provided by Madden and Tupes (1966). However, the regression formula for conversion of the AQE G scores need not actually be converted to RGLS for the present purposes, because it is only necessary to confirm the hypothesis of no difference across the experimental conditions for the AQE G scores. Tables 2-5 and 2-6 present the AQE G means, standard deviations, and t-values across experimental conditions. It was necessary to conduct two separate t-tests because each subject was not exposed to all of the test materials.

Table 2-5

AQE G Score Means, Standard Deviations, and t-values of Subjects in High and Low Factor Conditions for DMU, MMU, ESU, and CMR

	Mean	σ	t-value	Significance Level
SI Factor High	56.21	18.79	1.32	NS
SI Factor Low	61.76	14.80		

Table 2-6

AQE G Means, Standard Deviations, and t-values of Subjects in High and Low Factor Conditions for CMU, NMI, MFU, and NMS

	Mean	σ	t-value	Significance Level
SI Factor High	[illegible]	29.66	.97	NS
SI Factor Low	54.09	23.00		

The data presented in Tables 2-5 and 2-6 clearly confirm the hypothesis of no difference in general ability between the groups. In one instance, Table 2-5, the G score mean for the factor low group was higher, but the difference was not statistically significant. In the other instance, Table 2-6, the G score mean for the factor high group was higher, but this was also not statistically significant.

Intercorrelation Analysis

As will be recalled, each subject did not complete each readability/comprehensibility test. Thus, complete intercorrelation matrices could not be computed. Table 2-7 presents the product moment intercorrelation matrix for the DMU, MMU, ESW, and AFQT variables. The AFQT intercorrelations were included to determine the degree to which the readability/comprehensibility scores are related to intelligence as measured by the AFQT.

The intercorrelations as shown in Table 2-7 seem to be sufficiently low to indicate that the reading materials are relatively unique. Table 2-8 presents the i ercorrelation matrix for the same five readability/comprehensibility factors but in which the high load was involved.

Table 2-7

Intercorrelation Matrix for the Factor Low Condition for the DMU, MMU, ESU, CMR, and AFQT Variables.

	MMU	ESU	CMR	AFQT
DMU	.55	.27	.53	.36
MMU		.42	.28	.41
ESU			.29	.01
CMR				.31

Table 2-8

Intercorrelation Matrix for the Factor High Condition for the DMU, MMU, ESU, CMR, and AFQT Variables

	MMU	ESU	CRM	AFQT
DMU	.67	.67	.49	.54
MMU		.72	.71	.57
ESU			.71	.58
CMR				.43

The intercorrelations shown in Table 2-8 seem to indicate a substantial degree of dependence among the four readability/ comprehensibility test scores. For five of the six intercorrelations, 45 to 50 percent of the variance is common across variables.

Table 2-9 presents the intercorrelation matrix for the CMU, NMI, MFU, NMS, and AFQT variables for the low load condition.

As for the data presented in Table 2-7, the intercorrelations in Table 2-9 are low enough to indicate relative independence among the readability/comprehensibility materials. Finally, Table 2-10 presents the intercorrelation matrix for the same variables in the factor high condition.

Table 2-9

Intercorrelation Matrix for the Factor Low Condition for the CMU, NMI, MFU, NMS, and AFQT Variables

	NMI	MFU	NMS	AFQT
CMU	.37	.36	.48	.38
NMI		.00	.32	.34
MFU			.37	.34
NMS				.58

Table 2-10

Intercorrelation Matrix for the Factor High Condition for the CMU, NMI, MFU, NMS. and AFQT Variables

	NMI	MFU	NMS	AFQT
CMU	.30	.33	.44	.37
NMI		.17	.45	.57
MFU			.18	.22
NMS				.47

The data shown in Table 2-10 indicate that the textual readability/comprehensibility factors are relatively independent for this set of conditions.

In summation, the intercorrelational data seem to support the conclusion that most of the readability/comprehensibility factors are measuring relatively independent constructs.

Summary and Conclusions

The present chapter described the logic and structure of eight readability/comprehensibility constructs based on and derived from the Guilford structure-of intellect factors. It is held that these constructs possess advantages over previous measures of readability because the prior measures focus on structural aspects of the text while the new measures emphasize the cognitive involvement required for textual comprehension. Having derived and defined the readability/comprehensibility constructs, metrics reflecting their involvement in textual material were developed. The metrics were applied to samples of text prepared so as to reflect high and low loading on the new readability/comprehensibility measures. The textual materials were administered to a sample of Air Force personnel to determine whether or not statistically significant differences in comprehensibility were evidenced as a function of whether the materials were loaded high or low on the individual measures. The results indicated support for the following conclusions:

1. Reading material is more comprehensible to the extent that it deemphasizes the cognition of semantic units (CMU); i.e., as vocabulary diversity decreases.

2. The comprehension of prose is improved to the extent that reading material does not require the reader to form semantic (CMR) or word linkages. Comprehension is increased when these linkages are provided for the reader.

3. Only as much material as is necessary should be presented in figural materials (MFU). If the reader is not required to know or remember all details of a map or diagram, such details should not be presented to him.

4. Replication of facts (MMU) increases reading comprehension.

5. Reading material which provides semantic implications (NMI) or syllogisms will yield better comprehension than reading material which requires the reader to form semantic implications.

6. Material which provides mnemonic aids (NMS) will be more comprehensible than less organized material.

7. That material in which the reader is required to enumerate class members (DMU) will be less comprehensible than that material in which the class members are given to the reader.

8. Use of abbreviations and acronyms (ESU) has an especially disruptive influence on reading comprehension.

9. The metrics developed to measure the various readability/comprehensibility (SI) constructs may be employed to assess, at least partially, the readability of textual materials.

10. Each of the metrics developed for the SI constructs is reasonably independent from the others and from reading grade level.

CHAPTER III

PSYCHOLINGUISTIC DETERMINANTS OF READABILITY

Traditionally, the objective of readability researchers was to supply teachers with formulas which were simple enough for use without specialized training in linguistics and without complicated computation (Flesch, 1948; Dale & Chall, 1948). That is, traditional readability/comprehensibility research was conceived as being relevant to the measurement and evaluation of materials already written in suitable language. Bormuth (1969) indicated that the approach was entirely too narrow. The educator's problem, he says, is actually to transmit knowledge to students, using for the most part, language in written form, as the medium of communication. The effectiveness of the transmission processes, he said, can be increased by improving the student's ability to comprehend language and by controlling the difficulty of the language in which the transmission of knowledge is encoded. Controlling the difficulty of language communication can be accomplished by manipulating the language so as to make it less difficult. Bormuth (1969) asserted that modern researchers in the area must now regard readability research as being vital to the solution of every major aspect of the problem of increasing the effectiveness with which students organize the knowledge encoded in the language appearing in their instructional materials.

In his 1969 study, Bormuth examined many variables that correlated with reading difficulty. He stated that the principle implication of his studies is that it is urgent to undertake efforts toward a systematic analysis of the language comprehension process before it would be possible to design effective and systematic instructional materials; i.e., materials involving language that is suitable for student comprehension.

In order to provide such a technology, readability/comprehensibility research must be concerned not just with identifying variables which permit educators to predict difficulty of materials, but also with determining whether or not these variables can be manipulated. Specifically, the readability researcher should concern himself with identifying the manipulable linguistic variables which stand in causal relationship to difficulty. Bormuth (1969) contended that basic research had not, as yet, yielded data to allow the provision of such a technology. He also suggests that any technique employed to make material more readable/comprehensible would have to be general--applicable to persons at all levels of reading ability. Regardless of the reading aptitude of trainees, difficulty of comprehension, he said, is correlated with the same variables.

Although Bormuth himself indicated the available data to be too meager to produce a comprehensibility technology, his data, as well as the data of other investigators, suggest certain strategies to be used in attacking the readability/comprehensibility problem.

Sentence Depth

Yngve (1960) developed a model of sentence production which claimed that a person produces sentences by generating a "sentence structure tree" in a top to bottom-left to right direction. Accordingly, at any given time a speaker has produced only that portion of the left hand side of the tree necessary to produce the word spoken. As the speaker works down the tree, he produces both branches of a node, but he must store the right branch in memory while he is expanding the left branch. According to Yngve (1964):

It seems that as we speak, we incur commitments to finish our sentences in certain ways in order to make them grammatical.

As a string of words lengthens, such commitments must exist in a speaker's, listener's, and reader's memory if he is to complete the string in good grammatical form.

Suppose the sentence: "The new club members came early" were to be read. When a reader sees the word "The" he supposedly responds with the following two anticipations: (1) he expects to hear the rest of the noun phrase begun by "The," and (2) he also expects a predicate of some sort. "The," accordingly, is said to be embedded to a structural depth of 2. The next word "new" also has a depth of 2 because the reading of "new" elicits in the reader an expectation of completion of the noun phrase just as "The" did and affirms the already elicited expectation of a predicate. The word "club" also has a depth of 2 because the noun phrase still must be completed. The noun "members" has a depth of 1 because the only remaining commitment is the predicate. The verb "came" confirms the expectation of a predicate, but only partly because it, in turn, elicits an expectation of an adverb and hence is self embedded to a depth of 1. The adverb "early" is terminal (elicits no commitments) and therefore has a depth of 0.

The structural involvement, or the extent of embeddedness of each word in the sentence, can be characterized by the following set of numbers: 2, 2, 2, 1, 1, 0. These numbers also serve as an index of how much load the sentence supposedly imposes on memory. The mean of the set of these Yngve numbers may be taken as a measure of the structural complexity of the sentence as a whole. The greater the Yngve depth of a sentence, the greater its complexity in terms of structure.

We hypothesize that as structural complexity increases, readability/comprehensibility decreases. Bormuth (1969) found that sentence depth was correlated with the difficulty of a passage. Martin and Roberts (1966) held sentence length constant and varied the Yngve depth of sentences and found that sentences of lesser complexity were recalled significantly more frequently than sentences of greater structural complexity. For example, the sentence: "They were not prepared for rainy weather" has a mean depth of 1.29 (9/7= 1.29). It is recalled easier than the sentence: "Children are not allowed out after dark" which has a Yngve mean depth of 1.71 (12/7= 1.71). Wang (1970) has confirmed the finding that mean linguistic depth is a strong predictor of sentence comprehensibility.

These data indicate that readability/comprehensibility may be increased by either: (1) decreasing word depth within sentences, or (2) increasing the probability that the nodes in written sentences will effectively be stored in memory.

Decreasing Word Depth

It appears that the goal of decreasing word depth can be accomplished by deleting word modifiers whenever possible, and by breaking up long sentences into two--expressing action by one sentence and the meaning of the modifier by the other.

Ex: The^{2} very^{3} small^{2} boy^{1} rode^{1} the^{1} horse^{0}.
Mean depth= 10/7= 1.4

The^{2} boy^{2} rode^{1} the^{1} horse^{0}.
Mean depth= 6/5= 1.2

He^{1} was^{1} very^{1} small^{0}.
Mean depth= 3/4= .75

This approach not only reduces sentence depth but also increases readability by employing "referential repetition anaphora," the use of which Bormuth found was positively correlated with passage ease.

Increasing Probability of Node Retention

If it so happens that reducing sentence depth proves unfeasible, how can the probability be increased that the nodes of a sentence will be more effectively stored in memory? To accomplish this, the amount of information that must be held in memory at any one time must be reduced. As was indicated in Chapter II, Miller (1956) has shown that immediate memory (short term) is limited to processing only seven items (plus or minus two) at any one time. This suggests that if depth cannot be reduced, short sentences should be written. This should increase the readability/comprehensibility. Indeed, Foss and Crains (1970) have shown that the memory of words required of a listener or reader determines comprehension.

Morpheme Depth

Related to word depth within sentences is the problem of morpheme depth within words themselves. Bormuth (1969) speculated that the comprehensibility of an individual word may depend on how many morphemes are "buried" within it.

Ex: un/happi/ness

un = morpheme denoting "not"
happi = morpheme denoting a state of mood
ness = morpheme denoting a condition or quality

A person reading this word must have knowledge of the meaning of all three morphemes in order to comprehend the word. This suggests that the sentence: "The boy is sick because of his unhappiness" would be less readable/comprehensible than the sentence: "The boy is sick because he is not happy." It is probably because of this and similar issues that Bormuth found that sometimes longer sentences were more comprehensible than shorter ones.

Accordingly, it seems that not only a word depth index, but also a morpheme depth index would possess value for determining the comprehensibility of reading material. Words with a high morpheme depth index, more than likely increase reading difficulty.

Syllable Length

Bormuth (1969) found that although word length, measured in letters, provided an excellent index of difficulty, some of the most common words were long, not because they contained many morphemes, but because of the peculiarities of the English spelling system. This suggests that syllable length and word length (in letters) might be related to ease of comprehension. One who is interested in exploiting this conjecture might use a device to read a dictionary of the American language for the purpose of sorting out one and two syllable words. These words could then be used to compose a list (akin to a Dale [1931] list) of easily comprehensible words. If, indeed, minisyllabic and minimorphemic words are easier to read, then it would seem that a worthwhile endeavor would be to construct a minisyllabic and minimorphemic "thesaurus."

Structural Complexity of Sentences

Miller (1962) mentioned that the linguistic conceptions of the transformational grammarians (e.g., Chomsky, 1957) have important psychological implications. According to Miller, the main notion of such a grammar is the idea that the majority of sentences of a particular language are derived from a set of Kernel sentences by means of transformations. The passive sentence: "The girl was hit by the boy," the negative sentence: "The boy did not hit the girl," and the passive-negative sentence: "The girl was not hit by the boy" are held to derive, through transformations, from the kernal sentence: "The boy hit the girl."

Miller suggested that the analysis of a complex sentence into kernel sentences and transformations is useful as a model of language use. In support of this assertion, he cited two principal sources of evidence. First, he has found evidence which indicates that the difficulty of matching pairs of sentences from the same kernel is a direct function of the number of transformations separating them. Second, he points to a study by Mehler (1963), which showed that the majority of errors in the recall of complex sentences were syntactical errors; i. e., sentences which could be derived from the kernel sentence by applying one or more transformations. Miller interpreted this result to mean that when a complex sentence is heard, it is recorded and stored in memory as a kernal sentence along with a "footnote" indicating the necessary transformation. During recall of the sentence, recall of the kernal does not always result in recall of the "footnote." If this occurs, the syntactical error takes place.

Gough (1965), in an attempt to relate these findings to the problem of comprehensibility, assumed that people understand complex sentences only when they have been decoded to the underlying kernel sentences. If this is the case, he said, it follows that the latency of understanding a complex sentence should be a function of the number and nature of the transformations separating it from its kernel. He predicted that negative and passive sentences would be understood more slowly than kernels and that negative-passive sentences would be still more slowly comprehended. Gough's experimental manipulation of these variables indicated that: active sentences are understood faster than passive, affirmative sentences are comprehended faster than negative sentences, and passive-negative sentences are the slowest understood.

Slobin (1966) confirmed these findings when he asked his subjects to verify sentences of the four grammatical types--kernel, passive, negative, and passive-negative--with respect to pictures. He found the hierarchical order of comprehensibility to be: (1) kernels, (2) passives, (3) negatives, and (4) passive-negatives. But, when the sentences were made nonreversible; i. e., when the subject and object could not be interchanged as in "The boy drove the car," the differences in syntactic complexity between active and passive sentences "washed out." He suggested that the difficulty of understanding passive sentences may be partly attributable to the problem of keeping track of which noun is the actor. Fodor (1971) would concur. He indicated that the difficulty with passive sentences is not primarily due to the fact that they contain one more transformation than do actives. Rather, the passive voice destroys canonical phrase order (base structure) i.e., the decoding device which prefers to assume that the first noun phrase is a subject noun phrase is obstructed. He suggests two ways of increasing sentence complexity. One was to introduce into the sentence lexical items which are compatible with a relatively wide variety of deep structure types. The more types of deep structure a lexical item in a sentence is compatible with, the more alternative hypotheses the reader must entertain about the deep structure of the sentence. (See also Coleman, 1965.)

Fodor's second technique for increasing complexity is to eliminate or confound surface structure features which serve to "spell" the deep structure underlying the sentence. This suggestion is based on Fodor and Garrett's (1967) theory of sentence comprehension. It holds that a listener or reader constructs hypotheses about a sentence's underlying grammatical relations (deep structure) on the basis of cues in the sentence's superficial form (surface structure). Fodor and Garrett (1967) demonstrated that elimination of relative pronouns in center embedded sentences (sentences in which the subject and predicate are separated by a clause) appears to increase the difficulty subjects have in dealing with these structures. For example, sentence (1) below is predicted to be more difficult than sentence (2):

(1) The man the dog bit died.
(2) The man whom the dog bit died.

Hakes (1972) has essentially shown the same effect when "that" is deleted from a sentence; e.g., "John believed the girl was a fool."

Other factors that determine the structural complexity of a sentence, and thus comprehensibility, are the degree to which a sentence contains self-embedded structure and the degree to which its formation is right-branching or left-branching.

Schwartz et al., (1970) showed that as center-embeddedness increases (i.e., as clauses [from one to four] are embedded or added between subject and predicate) comprehensibility decreases.

Ex.: The angels that the theologians that the later cynics that modern science favors ridiculed counted stood on the head of a pin.

Schwartz et al.,(1970) also studied right-branching (where successive clauses are added to the right of the main clause) and left-branching (where successive clauses are added to the left of the main clause) sentences. An example of a right-branching sentence is: "The umpire called a balk that the southpaw pitcher hit that the coach replaced." An example of a left-branching sentence is: "The electricity powered toe chomping rock throwing lawn mower ran over the cord." They showed that increases in left-branching had no effect on comprehension but as right-branching increased, comprehension decreased.

Summary of Literature Indications

It seems that there are a number of psycholinguistic factors related to the readability/comprehensibility of textual materials. The data suggest a number of rules for making a sentence more comprehensible: (1) decrease word depth (Bormuth, 1969; Goss & Crains, 1970), (2) decrease morpheme depth (Bormuth, 1969), (3) change passive sentence to the active voice when there is a possibility of a reversal of subject and object; this reduces structural as well as semantical problems (Gough, 1965; Slobin, 1966; Fodor, 1971), (4) avoid center embedding whenever possible (Schwartz et al., 1970; Wang, 1970), (5) avoid right-branching sentences whenever possible (Schwartz et al., 1970), and (6) write affirmative sentences when possible (Gough, 1965; Slobin, 1966).

General Method

To determine whether or not the above listed factors affect the textual comprehensibility of reading matter, a compilation of reading materials reflecting the variables was developed and cast into a form appropriate for administration to USAF basic trainees at Lackland AFB. Texas.

These materials were placed into three booklets and administered to three groups of basic traineees (total N = 251) in three separate testing sessions. The time period of each of these sessions was approximately two hours. Whenever possible, all variations on each of the psycholinguistic factors were equally distributed throughout each of the three booklets. Administration was conducted by an Applied Psychological Services' psychologist who stood at the front of the room at a podium. At the request of the psychologist, the trainees were asked to open the booklet and to read, along with him, a passage on the purpose of the testing. In order to lessen any possible fear on the part of the trainees that the session was for "weeding out" purposes, they were told that their performance on the materials was a reflection of the difficulties of the materials themselves and not of their own intellectual and/or reading abilities.

After this, the recruits were given sample instructions for each type of test question to be found in the booklet: paragraph instruction, sentence instruction, arithmetic instruction, and picture verification instruction. After they had received these sample instructions, the recruits were told to get ready for data collection. The psychologist gave instructions to turn the page and read whatever material was on it. The trainees were given 0.5 seconds to read each word of the material (as did Coleman's, 1964, subjects); e.g., for a 10 word sentence, five seconds were allowed. When this time had elapsed, the psychologist said: "Stop; turn the page." Between each reading presentation and the testing on it, a page consisting of a column of six, one-digit numbers was presented. The recruits were told to add the column. This was incorporated to prevent any memorization from taking place (see Perfetti, 1969); ten seconds were allowed for performing this task. When this time had elapsed, the psychologist said: "Stop; turn the page." He then read to the trainees the instructions (which were also written on each of their own answer pages) as to how, properly, to respond to the previously read material. The response called for was one of four kinds: (a) 4, four option multiple choice questions on each of the paragraphs--60 seconds were allowed for response to each of the four questions, (b) an instruction to respond by writing the previously read sentence in full; for this task the subjects were allowed 30 seconds (see Wright, 1969), (c) fill-in questions, in which the trainees were asked a question about the previously read sentence and told to fill in the answer to it--here 10 seconds were allowed, and (d) a picture verification question, for which the subjects were asked to check off whether a picture at which they were looking was true or false as regards the previously read sentence--10 seconds were allowed for this decision and the response. At the end of the answer period, the psychologist said: "Stop; turn the page." He then read to the trainees the reading instructions which appeared on the next page. This procedure was continued until the entire booklet was completed.

Stimuli

The stimuli for the psycholinguistic factors which appeared in each of the three booklets in their diverse variations included 40 sentences of various Yngve depth (ranging from 1.09 to 3.51). Eight additional sentences were included in which Yngve depth was held constant while morpheme volume was varied. These sentences which varied in morpheme volume always contained the same number of words within each Yngve depth measure. There were two sentences with Yngve mean depth ($\bar{d}$) of 1.47. One of these had a morpheme depth (md) of 13; the other had a md of 17. Also, four sentences with $\bar{d}$ of 1.57 were included. Two of these varied md from 6 to 9 while the other two varied md from 9 to 13. There was another set of two sentences each with $\bar{d}$ of 1.82, but md being 10 in one and 17 in the other. The paradigm is shown in Table 3-1. A morpheme was defined as a unit of specific meaning (Coleman, 1971). It was hypothesized that sentences containing relatively more of these meaningful units will require more time to process. Additionally, four texts containing sentences in which Yngve depth was kept constant but in which morpheme volume was varied were included. The four texts had mean Yngve sentence depth as follows: 1.34, 1.43, 1.78, and 1.82. Within each of these texts, there were three paragraphs (containing exactly the same number of words) which varied in morpheme depth (or volume) from low through medium, to high.

Table 3-1
Paradigm for Sentences with Yngve $\bar{d}$ Held Constant with Morpheme Depth Varied (md)

	$\bar{d}$= 1.47	$\bar{d}$= 1.57A	$\bar{d}$= 1.57B	$\bar{d}$= 1.82
	md	md	md	md
1.	17	9	13	17
2.	13	6	9	10

The range of the number of morphemes in each of the paragraphs was as follows: low morpheme volume--42-69; medium morpheme volume--47-72; high morpheme volume--70-100. This is recapitulated in Table 3-2.

Table 3-2

Paradigm for Paragraphs with Yngve $\overline{d}$ Constant with Morpheme Volume Varied (mv)

ALL PARAGRAPHS CONTAIN SIX SENTENCES

Text $\overline{d}$= 1.34		Text $\overline{d}$= 1.43		Text $\overline{d}$= 1.78		Text $\overline{d}$= 1.62	
mv	words	mv	words	mv	words	mv	words
High	35	High	40	High	50	High	56
Med	35	Med	40	Med	50	Med	56
Low	35	Low	40	Low	50	Low	56

To determine the effects of syntactic complexity on comprehensibility, ten active, ten passive, and ten passive-negative sentences were derived from an original passive sentence having a Yngve mean depth equal to 1. 62. Responses to these sentences were of the true-false nature and, across booklets, the order of true and the false correct responses were switched. This paradigm is recapitulated in Table 3-3.

Table 3-3

Paradigm for Sentences Testing Transformational Complexity

Active		Passive		Passive-Negative	
True	False	True	False	True	False
1.	1.	1.	1.	1.	1.
2.	2.	2.	2.	2.	2.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
10	10	10.	10	10.	10.

To test the effects of embedding on readability/comprehensibility, ten sentences which were center embedded (from one to four clauses) were included along with ten of the same sentences in their deembedded form. There were also 10 left branching and 10 right branching sentences and 16 sentences which contained the complement "that." These 16 sentences were matched with 16 of the same sentences in which the complement was deleted. Table 3-4 summarizes the paradigm for these three effects. The sample N for each variable was approximately 251.

Table 3-4

Paradigm for That-Complement, Center Embedding, and Branching Sentences

Center Embedding		That-Complement	
Center Embedded	Deembedded	With "That"	Complement
1.	1.	1.	1.
2.	2.	2.	2.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
10.	10.	10.	10.

Branching

Left	Right
1.	1.
2.	2.
.	.
.	.
.	.
.	.
10.	10.

Results

Sentences Varying in Yngve Depth

Forty sentences of varying Yngve depth were constructed (ranging from 1.09 [low] to 3.5 [high]) and adapted whenever possible, to be of interest to a group of 18 to 20 year old Air Force recruits. For example, the sentence from Yngve, 1960: "When the very clearly projected pictures appeared the audience applauded." was adapted to read: "When the very well stacked broads appeared the recruits clapped." Sentences ranging from the lowest to the highest depth were distributed as equally as possible across the three booklets. The trainees were asked to respond to each sentence in one of three ways: writing out the sentence in full, answering a fill-in question, or verifying a picture representation of a sentence theme.

For 15 of the Yngve sentences, the subjects were asked to respond by writing all the sentences in full; for 16 of the sentences, they were asked to answer by a fill-in, and for 9 of the sentences, they were asked to respond by verifying a picture as being either true or false.

In scoring, when the sentence was to be written out in full, Perfetti's (1969) criterion was employed. This criterion states that an acceptable response is one in which the sentence is completely recalled with only inflectional errors at the bound morpheme level (e.g., omission of the past tense marker from a verb) allowed. In scoring the fill-in answers for the questions on a previously read Yngve depth sentence, a more lenient criterion was employed. Here, we reasoned that fill-in questions are ambiguous as to just what may be required for a proper answer. Thus, paraphrasing was allowed [see Hakes (1972) for a defense of this procedure].

For the picture verification answers, the subjects were presented with a picture above which was the caption: "Based on the picture, is the previous sentence true or false?" The subject was asked to respond by entering a check mark after either the word "true" or "false" which appeared below the caption. Scoring was performed dichotomously.

Table 3-5 presents the percentage of the subjects who performed correctly on each of the 40 sentences which were varied in Yngve depth.

Table 3-5

Percentage Correct for Each of the Sentences Varying in Yngve Depth (N = 251)

	$\bar{d}$	% correct		$\bar{d}$	% correct
1.	1.09	72	21.	2.17	100
2.	1.09	64	22.	2.17	100
3.	1.09	61	23.	2.17	100
4.	1.09	99	24.	2.18	31
5.	1.09	99	25.	2.18	100
6.	1.26	100	26.	2.27	78
7.	1.26	99	27.	2.27	0
8.	1.27	14	28.	2.29	100
9.	1.27	53	29.	2.31	0
10.	1.27	77	30.	2.31	67
11.	1.50	100	31.	2.40	99
12.	1.56	53	32.	2.40	99
13.	1.57	87	33.	2.35	0
14.	1.58	29	34.	2.56	76
15.	1.81	73	35.	2.59	100
16.	1.81	46	36.	2.60	5
17.	1.85	15	37.	3.15	1
18.	1.85	83	38.	3.26	100
19.	1.87	67	39.	3.44	5
20.	1.87	25	40.	3.43	36

Here, no consistent trend obtains in the progression from low to high depth. During the scoring, it seemed as though the extent to which an answer to a particular sentence was correct was dependent and sensitive to the measure that was used to obtain the answer; e.g. picture verification vs. writing out in full. Table 3-6 was therefore constructed to show the effect of the measure used on the probability of correctly answering each question. Table 3-6 suggests a trend toward the repeat in full measure yielding lower scores than the fill-in measure, which, in turn, appears to yield lower scores than the picture verification measure. A sign test, performed between the repeat in full measure percentage correct, and the fill-in percentage correct indicated no statistically significant difference in difficulty between these two measures. But, in a sign test performed between the fill-in and picture verification and the repeat in full and picture verification measures revealed differences in difficulty ($p < .01$).

Examination of the repeat in full measure column in Table 3-6 seems to suggest a trend toward greater difficulty as the sentences increased in Yngve depth. To determine whether this was indeed the case, these sentences were split at the $\bar{d} = 1.56$ level--those below this point were considered to be low in depth, those above it, high in depth. A Wilcoxon T-test was performed on these low and high depth sentences. The results indicated the sentences high in $\bar{d}$ were more likely to produce an incorrect answer on the repeat in full measure than those low in $\bar{d}$ ($p < .05$). The same procedure was employed for the fill-in measure and again a significant difference at the .05 level was obtained in incorrect answer production. Here, however, the sentences of high $\bar{d}$ were answered correctly, more often than those low in $\bar{d}$.

Table 3-6

Percentage Correct for Each of the Sentences Varying in Yngve Depth Based on Response Type

	$\bar{d}$	Repeat in Full %		$\bar{d}$	Fill in %		$\bar{d}$	Picture %
2.	1.09	64	1.	1.09	72			
4.	1.09	99	3.	1.09	41			
5.	1.09	99				6.	1.26	100
8.	1.27	14				7.	1.26	99
9.	1.27	53						
10.	1.27	77						
12.	1.56	77	11.	1.50	100	13.	1.57	87
14.	1.58	29	15.	1.81	73			
26.	2.27	78	16.	1.81	46	18.	1.85	83
27.	2.27	0	17.	1.85	15	21.	2.17	100
29.	2.31	0	19.	1.87	66	23.	2.17	100
33.	2.35	0	20.	1.87	25	31.	2.40	94
36.	2.60	5	22.	2.17	100	35.	2.59	100
37.	3.15	1	24.	2.18	31			
39.	3.44	5	25.	2.18	100			
			28.	2.25	100			
			30.	2.31	66			
			32.	2.40	16			
			34.	2.46	76			
			40.	3.43	36	38.	3.26	100

The small number of sentences using the <u>picture verification</u> measure did not allow employment of the above procedure, but examination of the data in this column suggests no difference in difficulty, as reflected by this measure.

Discussion of Yngve Depth Findings

In light of the prior data which appear to argue for a method bias, what do the results indicate concerning phrase structure analysis (Yngve depth) vis-a vis the issue of memory load/sentence recall and sentence readability/comprehensibility? First, the Wilcoxon T-test on the repeat in full measure indicated a likelihood of correct sentence recall to follow sentence mean depth in an inverse fashion. This finding corresponds with a finding of Martin and Roberts (1966), Mehler, (according to Martin & Roberts, 1966) and Bormuth (1969). To the extent that phrase structure analysis reflects recall, it does indeed seem that sentences of greater structural complexity (Yngve depth) impose a greater load on immediate memory than do those of a lesser complexity. Here, however, we are not concerned, primarily with recall of material as such, but rather with readability/comprehensibility--the generation (according to Fredriksen, 1973) of semantic information from linguistic inputs--of textual material. None of our subjects could correctly recall and write in full the sentence with Yngve $\overline{d}$ of 2.27 that follows: "Refusing to accept aid and comfort from the enemy, he planned to escape from camp." Yet, all of our subjects correctly responded by checking false to a picture of a man crossing a bridge and when they were asked to verify the picture against the sentence with the Yngve $\overline{d}$ of 3.26 that follows: "The news was bad, and he was depressed, so he jumped." Accordingly, the subjects demonstrated that they had indeed comprehended the deep structure of the sentence.

The role played by "surface structure" in the memory of sentences is presently a matter of some contention. The work of Martin and Roberts (1966) gives us, perhaps, the most workable hypothesis relating surface structure to sentence memory. They suggested that processing difficulty is a function of the number of left branches in a sentence. This hypothesis was derived from the phrase structure grammar put forth by Yngve (1960), and as indicated previously, is usually called the depth hypothesis. Other evidence supporting the depth hypothesis comes from the work of Roberts (1968) and Wang (1970). Our work also supports this hypothesis. The write in full measure data indicated the greater the $\bar{d}$ of the sentence, the more difficult sentence recall seemed to be. Our data, obtained from the fill-in and picture verification measures, however, failed to support the depth hypothesis. Other investigators who also have failed to find support for the hypothesis are: Perfetti (1968b, and 1969a, b). Rohrman (1968) and Wright (1969).

We are forced, therefore, to suggest that, while the Yngve depth measure may be useful in determining the load that a piece of reading matter may impose on immediate memory, its usefulness as a measure of the extent to which a sentence is readable/comprehensible is unresolved.

Sentences Varying in Morpheme Volume While Keeping Yngve Depth Constant

Eight sentences were included in which Yngve $\bar{d}$ was kept cor[illegible] ut in which morpheme volume varied. These sentences, whi[illegible] aried in morpheme volume, always contained the same number of words as their corresponding sentence within each Yngve depth measure. There were two sentences with $\bar{d}$ of 1.47, one with a morpheme volume of 13, the other with a morpheme volume of 17. There were four sentences with $\bar{d}$ of 1.57 but varying in morpheme volume from 6 to 9 for one set of two, but from 9 to 13 for the other set of two. There was another set of two sentences each with $\bar{d}$ of 1.82 but with morpheme volume being 10 in one and 19 in the other.

A morpheme was defined as a unit of meaning (i.e., content and function morphemes, a word, a base, plus inflectional and derivational affixes, such as prefixes and suffixes (e.g., disowned - three morphemes). It was hypothesized that sentences containing relatively more morphemes will require more central processing and, hence, yield lower scores. An example of a sentence that is low in morpheme volume is: "Happy and sad are opposite states." A sentence high in morpheme volume is: "Unhappiness and miserableness are depressive states." The measure used to test this variable was always the <u>write in full</u> measure.

To determine whether reading ability was a factor in determining comprehensibility for this morpheme volume measure, the subjects were split at the median of the total group of 251 subjects into high and low reading grade level (RGL) according to the regression equation given by Caylor et al., (1972). The equation allows prediction of RGL from AFQT score.

A 2 x 2 analysis of variance was performed on the percentage correct for all four Yngve depth sentences of high and low morpheme volume by high and low trainee RGL. No statistically significant effect of RGL or interaction of RGL with high and low morpheme volume was found. However, an F of 6.44 ($p < .01$) was indicated for morpheme volume. This confirms the hypothesis that as morpheme volume increases, comprehensibility decreases.

Because of the statistically significant effect of morpheme volume on comprehensibility indicated by the variance analysis, individual sign tests were performed on each sentence which varied in morpheme volume at each Yngve $\bar{d}$ level.

The directional hypothesis tested here was whether or not a sentence of a particular Yngve $\bar{d}$ is more difficult to process centrally if it is high in morpheme volume relative to one low in morpheme volume. Table 3-7 presents, for both the high and low RGL subjects, the results yielded by the sign tests for sentences in which the Yngve $\bar{d}$ was held constant but the morpheme volume was varied. For five of the eight comparisons in Table 3-7 there is a statistically significant difference in the predicted direction. Moreover, again, the possibility that differences were due to trainee RGL was not supported.

Table 3-7

Sign Test Results on Sentences Varying in Morpheme Volume (mv) with Yngve Depth ($\bar{d}$) Held Constant

	$\bar{d}$= 1.47	$\bar{d}$= 1.57A	$\bar{d}$= 1.57B	$\bar{d}$= 1.82
	mv 17, mv 13	mv 9, mv 6	mv 13, mv 9	mv 17, mv
High RGL	.001	NS	.001	NS
Low RGL	.008	NS	.018	.001

Paragraphs Varying in Morpheme Volume While Keeping Yngve Depth Constant

Four texts in paragraph form were also investigated. Each text contained three paragraphs and each paragraph contained six sentences. Each paragraph contained exactly the same number of words but, across paragraphs, morpheme volume was varied from low through medium to high. The mean Yngve depth values of the four texts were: 1.86, 1.78, 1.43, and 1.34. Morpheme volume was trichotomized over three levels: low, medium, and high.

The measure used to test the comprehensibility of these paragraphs was the correctness of response to multiple choice questions. For each paragraph within a text, the questions asked were always about the same word in the paragraph; e.g., the object of a particular preposition. An example of a paragraph that is low in morpheme volume is:

> The kid had just been hit when the mother came. The mother found one of the boys hiding under the table. The mother called the boy a dope. He felt the mother's anger. That was a sad thing. After the mother beat the child, he went straight to his room.

An example of a paragraph that is high in morpheme volume is:

> The double-agents had just been wiretapping when the agents appeared. The agents discovered one of the double-agents hiding during the encounter. The agents called the double-agent a co-conspirator. He outrightly denied the agents' accusations. This was an unbelievable state-of-affairs. After the agents transported the lawbreakers, they reported directly to the conventioneers.

An analysis of variance on the responses for each set of three paragraphs at the four levels of constant Yngve $\bar{d}$ by high and low trainee RGL indicated that RGL was a statistically significant factor in readability/comprehensibility on only the 1.43 and 1.34 Yngve depth paragraphs ($p < .01$). Morpheme volume was found to affect readability/comprehensibility on all but one of the Yngve depth texts (the text with a Yngve depth of 1.78). Morpheme volume proved to be a statistically significant readability/comprehensibility factor for all the other paragraph sets ($p < .01$). There was an interaction effect of RGL with morpheme volume for both the 1.86 and the 1.34 Yngve depth sets ($p < .01$).

Because of the statistically significant effects of morpheme volume indicated by the analysis of variance, individual sign tests were performed between each paragraph of a particular set of Yngve depth text; i.e., high against medium, medium against low, and high against low. Table 3-8 presents the results of these analyses (all two tailed tests). As indicated in Table 3-8, in only four of the 12 comparisons was the hypothesis that paragraphs high in morpheme volume are more difficult to comprehend than those which are low in morpheme volume not confirmed.

However, we note here a very recently published article. (Sherman, 1973) which suggested that sentences containing any negative components (be they either the word "not" or a prefix like "un") are harder to comprehend than are sentences not having these components. Examination of these texts here employed revealed them to contain confounding on this variable. For example, our high morpheme volume paragraph at the 1.4 Yngve depth level contained four negative components, as opposed to the medium morpheme volume paragraph at this level which contained only two negative components.

Table 3-8

Sign Test Results on Paragraphs Varying in Morpheme Volume (mv) with Yngve Depth ($\bar{d}$) Held Constant

Morpheme Volume Comparison	$\bar{d}$= 1.86	$\bar{d}$= 1.78	$\bar{d}$= 1.43	$\bar{d}$= 1.34
High-Med.	NS	.003-	.005	.003
Med.-Low	.005	.003	NS	.045
High-Low	.003	NS	.002	.003

This confounding, however, cannot account for most of the statistically significant results yielded by this investigation of the morpheme volume variable. Those paragraphs containing no negative components and those in which the negative components were equalized still showed that a paragraph high in morpheme volume is comprehended with greater difficulty than is one lower in morpheme volume.

Perfetti (1968b) found that lexical density (by which he means the ratio of content words to the total number of words in a sentence) was related to sentence retention. Perfetti concluded that: "... much of the memory space required by a sentence goes to the storage of semantic information carried by the lexical morphemes in the sentence..." (Perfetti, 1969b), and equally affects sentence retention and comprehensibility.

Coleman (1971, pp. 176-177) also reported that as the number of morphemes composing the words of a passage increased, comprehensibility decreased. He noted that anyone wishing to alter a passage to make it more comprehensible should reduce the number of morphemes and, thereby reduce the burden on central processes as well as the burden on visual processes.

Sentences Varying in Transformational Complexity

In an attempt to determine the effect of a syntactically complex sentence structure on readability/comprehensibility, sentences were constructed that varied in transformational complexity. Ten were active sentences (kernals); ten were passive sentences, and ten were passive-negative sentences. The active and the passive-negative sentences were derived by transformations of passive sentences all having Yngve $\bar{d} = 1.62$.

An example of active, passive, and passive-negative sentences respectively is: "They found the money lying in the corner." "The money was found lying in the corner." "The money was not found lying in the corner."

The method used to test the readability/comprehensibility of these sentences was always the picture verification technique. Each of the above sentences appeared twice in the booklets (but not for the same subject), once with its corresponding picture being true, and once with the corresponding picture being false. To discover whether reading ability was a factor determining sentence comprehensibility for these sentences, the subjects were split into high and low RGL trainee groups by means of the regression equation developed by Caylor et al., (1972).

Table 3-9 shows the p values obtained from sign tests between the active and passive sentences, the passive and passive-negative sentences, and the active and passive-negative sentences for the high and the low RGL subjects both when the pictures were true and when they were false. Table 3-9 indicates no differences in processing the active and the passive sentences whether or not the correct response was true or false or the RGL of the subjects was high or low. The high RGL subjects found the passive-negative sentences harder to process than passive sentences in seven out of the ten cases when the picture was true, but in only two out of ten cases when the picture was false. The low RGL subjects found the passive-negative sentences harder to process than the passive sentences eight out of ten cases when the picture was true, but in only four out of the ten sentences when the picture was false. The high RGL subjects found the passive-negative sentences harder to process than the active sentences in five out of the ten cases when the picture was true but in only two out of the ten cases when it was false. The low RGL subjects found the passive sentences harder to process than the active sentences in seven out of the ten cases when the picture was true but in only five out of the ten cases when the picture was false.

Table 3-9

Sign Test Results on Sentences Varying in Transformational Complexity

	TRUE			FALSE		
	ACT-PASS	PASS-PASS-NEG	ACT-PASS-NEG	ACT-PASS	PASS-PASS-NEG	ACT-PASS-NEG
	NS	.018	NS	NS	NS	NS
	NS	.016	.016	NS	NS	NS
	NS	.031	NS	NS	NS	NS
	NS	.002	.002	NS	.035	.090
High	NS	.090	.090	NS	NS	NS
RGL	NS	NS	NS	NS	NS	NS
	NS	.001	.001	NS	NS	NS
	NS	NS	NS	NS	NS	NS
	NS	NS	NS	NS	NS	NS
	NS	.001	NS	NS	.001	.002
% Significant	0	70	50	0	20	20
	NS	.016	.004	NS	NS	NS
	NS	NS	NS	NS	.004	.008
	NS	.008	.002	NS	NS	NS
	NS	.031	.033	NS	.020	.046
Low	NS	.020	.001	NS	NS	.046
RGL	NS	.031	.002	NS	NS	NS
	NS	.001	.001	NS	NS	NS
	NS	.060	NS	NS	.033	NS
	NS	.016	.011	NS	NS	.031
	NS	NS	NS	NS	.001	.001
% Significant	0	80	70	0	40	50

The observation of no differences in performance in processing simple active and passive sentences is not in accordance with certain prior findings; e. g., Coleman (1964, 1965), Mehler (1963), Gough (1965), and Slobin (1966), using as did we, the picture verification task. On the other hand, Slobin (1966) found that under some conditions passive sentences were harder to process than active sentences. But when these sentences were made nonreversible, that is, when the subject and the object could not logically be interchanged as in: "The horse was seen running around the track" the differences in syntactic complexity "washed out." That is, both were equally comprehensible. Other investigators who have failed to find comprehensibility differences between active and passive sentences (and thus failed to support the transformational grammar model) are Martin and Roberts (1966), Perfetti (1969), and Moore and Biederman (1973).

When Slobin (1966) found that making his sentences nonreversible resulted in a "wash out" in the differences in complexity between active and passive sentences, he suggested that the difficulty in understanding passive sentences may be partly attributable to the problem of keeping track of which noun is the actor. Fodor (1971) would agree. He believes that the difficulty with passive sentences is caused, not primarily by the fact that they contain one or more transformations, but because the passive voice destroys canonical phrase order (base structure)--the decoding device which prefers to assume that the first noun phrase is a subject noun (Fodor, 1971, p. 125). He went on to suggest two ways of increasing sentence complexity: (1) introduce into the sentence lexical items which are compatible with a relatively wide variety of deep structure types [the more types of deep structure a lexical item in a sentence is compatible with, the more alternative hypotheses a reader must entertain about the deep structure of the sentence (see also Coleman, 1965)], and (2) eliminate or confound features of the surface structure which help to "spell" the underlying deep structure of the sentence.

Fodor based the second suggestion on Fodor and Garrett's (1967) theory of sentence comprehension which holds that a listener (and presumably a reader) constructs hypotheses about the underlying grammatical relations (deep structure) of a sentence.

Presumably, this is what causes the difference in comprehensibility between active and passive sentences when their subjects and objects are reversible. An examination of the sentences used in our experiment reveals the subjects and the objects to be essentially non-reversible. This fact would seem to account for our not finding differences in comprehensibility between our active and passive sentences.

Table 3-9, however, indicates that the low RGL subjects found the passive-negative sentences harder to process than the passive sentences in eight of the ten cases when the picture was true, but in only four out of the ten cases when the picture was false. The high RGL subjects found the passive-negative sentences harder to process than the active sentences in five out of the ten cases when the picture was true but in only two out of the ten cases when it was false.

In these cases we have replicated the results of many prior studies which indicated that passive negatives are more difficult to process than are either active or passive sentences; e.g., Gough (1965).

We seem to have shown in terms of Savin and Perchenoch's (1965) interpretation, that kernal sentences occupy less space than do passive-negative sentences; and also, apparently, passive sentences. We have also replicated Slobin's (1966) finding concerning the interacting effects of truth and falsity on the picture verification task with the transformational variables. He found, as did we, that when the picture was true, more errors were made to the passive and passive negative sentences than when the picture was false. He found this to be the case with subjects ranging from ages six through twenty. We, however, noted this to be the case more so for our low RGL subjects than for our high RGL subjects. Wason (1959, 1962) Eifermann (1961), McMahon (1963), and Gough (1965) all reported that their subjects' behavior reflected a greater difficulty when dealing with true negative statements than they did with false negative statements. Slobin (1966) reported that several of his youngest subjects refused to accept any of the negative sentences as being true. He suggested that perhaps this interaction between truth and affirmation can be accounted for in terms of an "atmospheric effect." He postulated that "affirmation" is the language of truth and that negation is the language of falsity (Slobin, 1966). There may be a tendency to call affirmative sentences true and negative sentences false. In Slobin's experiments, as in ours, a negative was true because it described the reverse of the picture. Because this is the case, Slobin went on to suggest that true negatives are more difficult to verify that false negatives when the following conditions are present: (1) pictures are used as referents, and (2) both types of sentences are evaluated in regard to the same constellation of actors and action. Stated in other terms, condition 2 requires that the sentence and the picture have the same content (i.e., the same noun and verb). These conditions were present in ours and in Slobin's study: true passive-negative sentences tended to be more difficult to verify than did false passive-negative sentences, as also were the true passive sentences, as compared to the false passive sentences. The subject of false affirmative and true negative sentences does not correspond to the actor in the picture, but in the case of true affirmative and false negative sentences this correspondence does obtain. Such a "mismatch," as Slobin calls it, may pose problems to a subject if part of his "strategy" is to match the stimulus sentence by generating a true affirmative sentence describing the picture. This problem of "mismatch" accounts fairly well for the difficulty in dealing with passive and passive-negative sentences.

Perhaps a more elegant way of describing the process discussed above is seen in Clark and Chases' (1972) "Model A" of a theory of sentence-picture comparison. Their theory of sentence-picture comparison (verification) was designed to account mainly for a limited type of sentence verification task. Here, a subject is shown a display containing a sentence like: "Star isn't below line" and a picture of, say, a star above a line. The subject is asked to read the sentence, look at the picture, and indicate as quickly as possible whether the sentence is true or false. The sentences used in this task always made use of above or below and described the vertical position of two geometrical figures. [Although the theory Clark & Chase presented is meant primarily to account for the response latencies of their subjects in dealing with the above tasks, it is applicable to and can also account for erroneous responses.] Because their theory deals in a great part with the verification of negative sentences, part of it can be traced to extensive earlier work on negation by Wasen (1961), Eiferman (1961), Gough (1965, 1966), Slobin (1966), and others. Clark and Chase acknowledge that Trabasso (1970) and Trabasso, Rollins and Shaughness (1971) have independently formalized almost the identical general model for the comprehension of negation.

In "Model A," Clark and Chase (1972) divided the sentence picture comparison process into four identifiable stages. At Stage 1, the subject is said to form a mental representation of the sentence. At Stage 2, he forms a mental representation of the picture. At Stage 3, he compares the two representations. At Stage 4, he makes a response. This model is capable of predicting the time it will take a subject to verify a particular sentence and assumes that the time for each separate process is additive. Experiments verified that the model receives excellent support in the terms of verification times and in percentage of errors made to each kind of sentence. Their data are also consistent with ours in showing that more errors are made to "true" negative pictures than to "false" negative pictures. With our subjects, though, we saw a tendency for the low RGL people to be more susceptible to errors. It would thus appear, that when writing, to ensure readability/comprehensibility, the use of the passive-negative voice should be avoided and, especially, this practice should be followed when writing for those with low reading grade level.

Sentences Varying on Other Structural-Complexity Dimensions: Complement Deletion, Center Embedding, Left and Right Branching

Fodor (1971) suggested that one way to increase sentence complexity would be to eliminate or confound surface structure features which serve to "spell" the deep structure underlying the sentence. He based his assertion on Fodor and Garrett's (1967) theory of sentence comprehension. This theory holds that a listener or reader constructs hypotheses about a sentence's underlying grammatical relations (deep structure on the basis of cues in the sentence's superficial form (surface structure). Fodor and Garrett (1967) have shown that elimination of relative pronouns in center embedded sentences (in which the subject and predicate are separated by a clause) appears to increase the difficulty of dealing with these structures. The deletion of the relative pronoun "whom" from the sentence: "The man the dog bit died," supposedly increases the ambiguity of the sentence. Accordingly, its comprehensibility is decreased.

Hakes (1972) has essentially shown the same effect when "that" is deleted from a sentence such as: "John believed the girl was a fool."

Other factors that determine the structural complexity of a sentence and thus its comprehensibility are the extent to which the sentence contains self embedded structures and the degree to which its formation is left or right branching.

Schwartz et al.,(1970) have shown that, as center embeddedness increases (that is, as clauses are embedded or added [from one to four] between subject and predicate), comprehensibility decreases. Wang's (1970) data supported this finding.

Schwartz et al., (1970) also studied right branching (where successive clauses are added to the right of the main clause) as in the sentence: "The umpire called a balk that the southpaw pitcher hit that the coach replaced," and left branching (where successive clauses are added to the left of the main clause as in the sentence: "The electricity powered toe chomping rock throwing lawn mower ran over its own cord." They demonstrated that increases in left branching had no effect on comprehension but as right branching increased, comprehension decreased.

The methods and results of an examination of the role of each of the above sentence complexity factors relative to readability/comprehensibility are presented below.

Complement Deletion

Sixteen sentences which contained the complement "that" and 16 sentences in which this complement was deleted were included in the data collection booklet. The measure used for answering these sentences was always the <u>fill-in</u> measure. Because we reasoned that fill-in questions are, to some extent, ambiguous as to just what may be required for a proper answer, we scored a paraphrase as a correct answer (see Hakes, 1972, for a defense of this procedure). Subjects were again split into high and low RGL trainee groups for purposes of analysis.

A 2 x 2 analysis of variance of percentage correct for complement present or absent and trainee RGL "high" or "low" indicated no statistically significant main or interaction effects. More errors were noted on five of the sentences not containing the complement; there were fewer errors in nine of the sentences not containing the complement, and in one of the sentences the errors made both with and without the complement were equal.

Center Embedding

Ten sentences which were center embedded (containing from one to five clauses) were constructed. These were matched with ten sentences in their deembedded form. An example of a sentence center embedded by five clauses is: "The dragon, giving no evidence of surrendering under the numerous attacks of the knights who charged at him with a loud clash of swords, was forcing them to retreat" (from Wang, 1970). Deembedding this sentence yields: "The dragon was forcing the knights to retreat because he showed no evidence of surrendering under their numerous attacks when they charged him with a loud clash of swords." Again, the fill-in measure was employed (allowing paraphrasing) as the response mode.

Table 3-10 presents results of sign tests performed on each of the ten embedded and deembedded sentences relative to the hypothesis that the deembedded forms are more readable/comprehensible.

It can be seen that in seven out of ten cases, there was no statistically significant difference in responses to the sentences in either their embedded or their deembedded form. However, in three of the cases, statistically significant differences in favor of the hypothesis that embedded sentences are less comprehensible were obtained.

Table 3-10

Sign Test Results between Ten Embedded and Ten Deembedded Sentences

Sentence	Test Results
1	NS
2	.001*
3	.002*
4	NS
5	NS
6	NS
7	NS
8	NS
9	NS
10	.001*

*in direction of hypothesis

Right and Left Branching

Ten sentences with four clauses to the left of the main clause and ten sentences with four clauses to the right of the main clause were constructed and employed as stimuli to test the hypothesis that left branching sentences are more readable/comprehensible than are right branching sentences. Again, the fill-in response mode (allowing paraphrase) was employed.

Table 3-11 presents the results of sign tests performed on each of the ten right and ten left branching sentences relative to the hypothesis that the left branching sentences are more readable/comprehensible than right branching sentences. Here, it can be seen that in one case there was no statistically significant difference obtained between the right and left branching sentence; in another case there was a significant difference in favor of the hypothesis. But, in the eight remaining cases, significant differences were noted in the wrong direction.

Table 3-11

Sign Test Results Performed between Ten Right and Ten Left Branching Sentences

Sentence No.	p Value
1	.0003**
2	.0003**
3	NS
4	.0003**
5	.004**
6	.01**
7	.001**
8	.0003**
9	.007*
10	.0003**

*in favor of hypothesis
**against hypothesis

Discussion

These results are interpreted as failing to show that deletion of the complement "that" caused a loss of sentence comprehensibility for our subjects. Our subjects, contrary to some of the findings reported in the literature, found it easier to comprehend right branching sentences than left branching sentences. But, at least marginally, embedded sentences were less comprehensible than the deembedded sentences. What possible reason might there have been for the present findings?

The stimulus sentence for the "that" complement deletion factor were variations of those used in the Hakes (1972) experiment (i.e., all were double self embedded with the complementing verb always the verb of the sentence's main, independent clause). Our sentences were read by subjects; Hakes' subjects heard the sentences. Our subjects were asked to answer a question concerning the sentence; Hakes' subjects were asked to paraphrase the sentence after performing a phoneme monitoring task. These differences in tasks may account for the difference in findings across the two studies. Additionally, the present results may be due to the fact that the effect of deleting the "that" complement seems "weak." Hake (1972) found that the results of the monitoring task strongly supported the hypothesis that deletion of the "that" complement increases comprehension difficulty; however, the results of the paraphrasing task did so only weakly. Our fill-in response mode is more closely related to a paraphrasing task than to a phoneme monitoring task.

The findings of the present study, relative to the deembedding of sentences, were often in the proper direction--although statistically significant results were not obtained.

One, two, or four subordinate clauses were generally easier to comprehend in their deembedded form. This was not, however, the case for some sentences with three subordinate clauses. We note also that in the prior studies relative to this variable (Schwartz et al., 1970; Wang, 1970; Hamilton & Deese, 1971) the subjects heard the sentences and were asked to express their degree of judgment as to the sentences' comprehensibility on a scale. Our subjects, on the other hand, read the sentences and were asked questions about them.

The finding that our subjects comprehended right branching sentences more readily than left branching sentences is again believed to reflect data collection method sensitivity. Additionally, we note that Hamilton and Deese (1971) found that right branching sentences are more readily comprehended than are center embedded sentences. They attribute this finding to the fact that in the right branching sentences the subject and predicate of each clause occurs contiguously. Contiguity of grammatical structure may represent an explanatory construct in this regard.

The results reported above have demonstrated that a number of psycholinguistic variables affect readability/comprehensibility. These factors are transformational complexity (specifically, passive-negative sentence difficulty), morpheme volume, and (marginally) the structural complexity factor of center-embeddedness. This study was unable to replicate certain findings from other research; specifically that: (a) passive sentences are more difficult to comprehend than are active sentences, (b) deletion of the "that" complement causes incomprehensibility, and (c) right branching sentences are less comprehensible than are left branching sentences.

The results indicated that the Yngve depth factor, while important, was measure-sensitive and probably related more to short term memory load than to comprehensibility, per se.

Except in the passive-negative sentence case, trainee reading grade level was not a particularly significant factor here; varying these psycholinguistic factors had, for the most part, equivalent effects on readability/comprehensibility for both high and low reading grade level subjects. It seems, on the bases of these research findings, that methods whereby readability and comprehensibility may be increased by a writer of textual material have been identified. These findings represent an initial attempt at determining psycholinguistic aspects of readability/comprehensibility and further similar and related research is needed in order to establish an adequate technology of written instruction.

CHAPTER IV

FEASIBILITY OF AUTOMATIC CALCULATION OF READABILITY/COMPREHENSIBILITY METRICS

The purpose of this chapter is to present views on the extent of the feasibility of calculating text comprehensibility measures automatically. The utility of several such measures was presented and discussed in prior chapters of this report and shown to be reasonable approaches to the scientific measurement of comprehensibility. These same measures are here examined with respect to approaches which could be taken to computerize their determination.

First, a background review is presented to give the reader a summary of the state-of-the-art in the field of automatic text processing, now called semantic information processing. Then, the measures to be mechanized are presented, together with possible approaches for accomplishing mechanization. The names of the specific measures, together with the level of difficulty, for automation are summarized in Table 4-1.

The Future of Semantic Processing

It is interesting to conjecture about the future in this field. The extent to which automation is determined to be feasible (together with later success in its implementation) could have a far reaching effect on text preparation and eventually on writing styles. A rapid increase in the routine operational use of computers to prepare text for publication is now being experienced. A 1971 survey of available on-line editing systems included about a dozen computer programs called "text editors" (Van Dan & Rice, 1971). Recent developments have extended this trend, and it is expected to continue. Within a decade, it is believed that a significant percentage of all published material from newspapers to encyclopedias will be computer processed.

Table 4-1

Summary of Comprehensibility Measures and Their Difficulty for Automation

Structure-of-Intellect	Likelihood of Success in Automatic Computation
1. type token ratio	1. relatively simple
2. cognition of semantic relations	2. difficult but possible
3. memory of semantic units	3. relatively simple
4. evaluation of symbolic implications	4. relatively simple
5. cognition of figural units	5. simple (initial approach)
6. convergent production of semantic systems	6. simple
7. convergent production of semantic implications	7. needs further study
8. divergent production of semantic units	8. needs further study
Psycholinguistic	
1. Yngve depth	1. difficult but possible
2. morpheme depth	2. relatively simple
3. transformational complexity	3. relatively simple
4. self embedding	4. relatively simple
5. left/right branching	5. difficult
6. effect of deleting complements	6. relatively simple

114

It is obvious that part of the computer processing of text for publication is the "typing" or optical reading of the text in machine readable media. Some of the more advanced text editing programs also use a computer-stored dictionary. With this capability, a computer program can accomplish functions such as automatic hyphenization, page numbering, indexing, page layout, spelling checks, centering of headings and the like. We believe that, some time in the future, there will be a natural extension of this type of computer processing so as to add the capability of determining one or more comprehensibility indices. Accordingly, the project, to the extent that it is recommended here as feasible, could result in a programming logic flow for text processing which, in turn, could become the "back end" of more routine text handling procedures now available or being developed.

As a last introductory thought, we note that the results presented apply exclusively to the English language, as would any computer technique resulting therefrom.

State-of-the-Art

Like most fields of endeavor, the handling of natural language text has benefited substantially from the availability, within the last two or three decades, of automatic data processing systems.

Sedelow (1970), in a discussion of the use of computers in the humanities, confirms that tasks such as automation of text analysis is now very much in the field of interest of the humanist. He writes that:

> Tasks such as syntactical analysis, stylistic analysis, structural analysis, etc., are of interest in traditional humanistic studies and are vital to computer-assisted instruction, automatic abstracting, information retrieval, machine translation, and the analysis and synthesis of graphics.

Word Frequency Applications

Automatic preparation of concordances by computer is one of the earlier applications of the computer to text processing. As identified by Bowles (1967), concordances of the Bible and the Dea Sea Scrolls were published as early as 1957. And, it may be inferred from the survey conducted by Sedelow (1970), that this technique had already become routine and commonplace, with concordances available for such varied texts as the poems of Matthew Arnold, W. B. Yeats, and Emily Dickenson, along with the writings of William Blake and Lord Byron. The ACM Computer Programs Directionary (Faden, 1971) describes a FORTRAN IV program used in preparing a concordance analysis of the plays of Eugene O'Neill. Mcst recently, after 25 years of data collection and analysis, a concordance of 179 works (mostly attributable to St. Thomas Acquinas) covering ten million words was completed. In summary, Parrish observes (Bowles, 1967):

> The successful completion of a computer concordance makes the making of concordances by hand old fashioned, obsolete. The making of dictionaries of larger bibliographies by hand will soon enough in the same way become obsolete.

A similar type of application is that of the Key Word In Context (KWIK) index designed by Luhn (1969). This index places the word of interest in the center of a single print line and provides as much of the context in which the word is embedded as the print line will hold. It is therefore both an abstracting and indexing technique. The KWIK technique is now used routinely in indexing periodicals and the like. Although the technique is applied mostly to indexing scientific materials, it is generally applicable to indexing of any text. The KWIK index is an example of one useful system which relies on cross referencing titles by all key words in the title. Other approaches, by selecting words which occur in a document more frequently than normal usage would predict, generate a set of content words which is suitable not only for indexing but also for abstracting and later information or document retrieval.

The ability of the computer to count word frequencies--a by-product of concordance generation--led to the insinuation of data processing techniques into stylistic analysis and "attribution" studies--the determination of the authorship of a given work. One early effort to resolve a question of "real" authorship was that reported by Mosteller and Wallace (1963). In this case, it was concluded that James Madison (not Alexander Hamilton) wrote The Federalist Papers. The analysis was performed on the basis of about 100, 000 meaningful words using statistical techniques and an electronic digital computer.

Since that time, the computer has been employed for interesting and varied attribution tasks, including determining that the Illiad had only one author and that the book of Isaiah had two distinct authors.

Sedelow (1970) reported that:

> Humanists are becoming increasingly interested in using the computer to explore relationships among the words and other linguistic units and among words and textual units, as well as relationships among categories describing behavior of words. These categories include the syntactic, semantic, temporal, and spatial.

She also reported briefly on a "General Inquirer" computer program for content analysis. The program looks toward having some conceptual or theoretical relationship which is specified in advance by a research scholar. As described by its authors (Stone et al., 1966), the program has been used to study folktale themes and in distinguishing "genuine" vs. "pseudo" suicide notes.

Another analytic method, initiated by Sedelow, is the Verbally Indexed Association Program, which looks for words in relationship to their frequency of occurrence. Its purpose is to reveal structuring concepts, themes, or attitudes in a text. The program has been used to examine prose, historical writings, and political campaign speeches.

117

Another example of computer analysis on the basis of word frequency determination is that proposed by Johnson (undated). Johnson used computer methods to facilitate new language learning via the reading process. The technique results in computer printed text in which only relatively rare words that the student should not learn at his current level are identified by translating them in the margin. Other words are marked (and translated) to indicate that they should be mastered on the first occurrence. Word selection is based on actual frequency of occurrence and a preselected number of words to be learned each year. This defers learning of less frequently used items without burdening the student to do the selection. His attention is focused exclusively on those vocabulary items that are the most significant for him at his particular learning level, ignoring less important words.

Dictionary Development

There is now a growing availability of word lists and dictionaries in magnetic tape form for computer aided applications. Several sources of such materials exist such as Brown University's 1, 014, 312 words of running present-day American text (Francis, 1964). The Semantic Foundation project (formerly Systems Development Corporation Lexicography project) offers magnetic tape transcripts of Webster's <u>Seventh New Collegiate Dictionary</u> and the <u>New Merriam Webster Pocket Dictionary</u> (Reichert, Olney, & Paris, 1969). Dozens of users of these dictionaries are reported (Olney & Ramsey, 1972). The availability of such material would simplify research and experimentation with, or operational use of, the several readability/comprehensibility techniques which require such aids.

Natural Language Inquiry Systems

Our goal in this current work is to mechanize, via computer, the analysis of sentence structure so as to handle the logic and calculation sequences required to determine the selected comprehensibility measures described in prior sections of this report. Work based more directly on this type of requirement has not been altogether lacking. In recent years, the technology has made important strides. The main impetus of this progress has been principally the desire to have computers respond to questions posed in English. In response to this need, various workers have been active. Table 4-2 cites a variety of early developments extracted from Green (1963).

Table 4-2

Some Early English Language Processing (Inquiry) Programs

Author	Date	Program Name	Function
McCarthy	1959	Advice Taken	draws conclusions from English statements
Lindsay	1960	Family Tree	abstracted into familial relationships and generate family tree
Phillips	1960	Responder	answers questions in English about pre-stored sentences
Simmons, et al	1961	Synthex	like Phillips in a larger context
Green, et al	1960	Baseball	answers simple questions about baseball in English, extracted from data in a data base

119

It is note that Phillips' program uses a stored dictionary in order to accomplish a syntactic analysis of sentences. In the "Baseball" program, questions are also syntactically analyzed, and the missing questioned information is sought in a suitable prepared data base.

In another early work, Householder (1961) reported on the development of a general mechanical routine for the reduction of complex sentences to their constituent simple sentences without loss of information content. Secondarily, he worked toward an artificial language (based on English) suitable for storage, translation, or manipulation.

More recently, there has been additional and substantial work in the field of natural language inquiry systems closely related to the task at hand. This is seen as a very positive influence on the probability of success of automating comprehensibility. Natural language inquiry systems are based on new computer data base storage and retrieval techniques developed in the late 60's and early 70's. At least, the following five well-recognized groups are engaged in the development of the capability to accept input queries to a computer data base in English rather than an artificial inquiry language (though only a limited English subset [grammar] is, of course, permitted):

Systems Development Corporation
California Institute of Technology
Bolt, Beranek, & Newman, Inc.
Massachusetts Institute of Technology
University of Texas

An illustration of a technique applied in developing such a question-answering machine is given by Simmons in Borko (1962). These developments, in turn, are spurred by the facts that: (1) remote access to data bases is becoming much more common, and (2) more worthwhile data bases are becoming increasingly available --even on a commercial basis. These trends are expected to continue with the end result that the field of computational linguistics will be an important, if not critical, research and development area for at least the balance of the 70's.

120

These natural language inquiry systems each utilize a parsing program which accepts the (restored) natural language inquiries and determines the semantic interpretations of these inquiries, translating them into expressions which generate various actions on the data bases. It is expected that this specific experience in the parsing of English would be of direct assistance in the development of automatic techniques for readability/comprehensibility measurement.

Currently, developers report times from less than .10 to 20 seconds, depending on approach and the end use of the parse result, to automatically parse an English language inquiry (many nuances are not admitted) including flagging of some grammatical errors.

A related work is the parse-a-system program for syntactical analyses of English text for the IBM 7094 (Faden, 1971). Here, the program inputs grammar coded English text, one sentence at a time (using parsing logic to select grammar code in pairs or adjacent constituents), and presents each pair to a table of previously input grammar rules for comparison.

Machine Translation

One of the earliest serious attempts to use computers for semantic applications was the machine translation experiments started in the early 1950's. Despite substantial funding, automatic or semiautomatic translation between languages with the aid of a computer was beset by ambiguity problems, and early optimism soon degraded. It is now generally agreed that machine translation is still a technique which will not yield text of sufficient quality to be of practical use. Minsky (1968) summarized the situation:

> The poor results in early translation attempts resulted from the hope that adequate syntactic analyses of sentences could be made without an apparatus for assessing the plausibility of proposed meanings. This gamble didn't pay off. It is now apparent that the meanings must be taken into account to resolve ambiguities even within coherent discourse in a single language, let alone in translating. One needs methods for representing the entities being discussed and the relation between them, as well as enough logical inference capacity to make common sense deductions about the consequence of these relations. (underlining added)

Simmins (Borko, 1962) had earlier concluded that the problems associated with the actual implementation of machine translations are more apparent than are the solutions: "There exist problems recognizing parts of speech, of workers syntactic analyses, of logical inference on the basis of syntactic and semantic structures, and a myriad of problems concerned with the meaning of words and sentences."

Accordingly, those engrossed in machine translation made their best contribution to semantic processing by formalizing the difficulties involved and partially, as a result, substantial stimulus was given to linguistic research projects.

Summary of Literature Indications

The technological developments of recent years, accordingly, point to the practical feasibility of automating the determination of some readability/comprehensibility measures for prose English text. Several developments have combined to bring about this favorable situation.

1. data processing technology, equipment, and software languages have become available over the years

2. extensive basic research has been carried out in the important fields of linguistic (grammatical) parsing techniques and syntactical analyses

3. an increasing number of computer applications dealing with processing of English words have been successful on projects such as developing concordances, author attribution studies, text editing, and English language inquiry systems

4. utilization of computers for text editing of newspaper and book publications has become routine

5. larger scale operational usage of optical scanning equipment for test reading is currently available

6. variety of English language dictionaries are available in magnetic tape form

As a result of these developments, considerable optimism has developed relative to the practicability of automating the calculation of several of the more mechanical readability/comprehensibility measures described earlier.

Explanation of possible approaches to this automation and identification of specific measures for first automation constitute the remaining sections of this chapter.

Manually Determined Indices

At the onset, we note that a long list of readability/comprehensibility measures has been offered for consideration over the past 30 years. A sample of those considered to be of principal interest is contained in Table 4-3. For convenience, they have been grouped into three classes: structure complexity, word divergency, and parts of speech. These deal principally with what one might call mechanically oriented factors. They deal with quantities of words, sentences, syllables and their occurrences, but are not concerned with meanings of words or phrases per se. They have been in use for some time not only because they could measure reading difficulty in some sense, but also because they were suitable to relatively easy calculation by hand. A comprehensive summary of these techniques is presented in Williams, Siegel, and Burkett (1973). These measures have been used principally to determine the reading grade level of text.

Table 4-3

"Classical" Readability Measures

Readability Measure	Authors or Developers
Complexity of Structure	
letters per word	Gray/Leary, Lorge, Bormuth
words per sentence	Flesch, Spache
vowels per word	Coke/Rothkopf
syllables per word	Flesch, McLaughlin
Word Divergency	
different words	Vogel/Washborne
words in Thorndike's list of 10,000	Vogel/Washborne, Ojeman, Bormuth
words not in Dale's list of 3,000	Dale/Chall
words not in Dale's list of 3,000	Spache, Gray/Leary, Lorge
words not understood	Jacobson, Dale/Tyler
Parts of Speech	
prepositions	Lorge, Ojeman, Szalay
pronouns	Szalay
infinitives	Ojeman

The limitations of these measures and the advantages of their automation by computer program, along with a brief description of the program, were reported by Jacobson and MacDougall (1970):

> Most readability formulae can be criticized for depending on word lists which are out of date. In addition, readability formulae were ready-made for use by laymen and other non-computationally oriented persons, resulting in restrictions on the clerical effort and computational skill needed to apply the formula. Samples of textual material rather than entire texts were used. These samples were often inadequate and not representative of the materials from which they were taken. In using such samples, counts were made of variables which measure readability. Such variables were sentence length, word size, word difficulty (as measured by word lists), and number of syllables, etc. Most formulae were limited to two or three variables made on samples of one or two thousand words. Both limitations were necessary because man, not a machine, was doing the work.
>
> The automated feature and related analysis offer specific advantages to the production of programmed materials in two principal ways: first, directly, in the writing, revision and evaluation of materials, through experimentation, in program definition and evaluation of the relative influence of methods on program structure.
>
> The second principal advantage of the automated analysis is that it offers a promising approach to the definition and evaluation of programmed materials, the identification of significant frames, response and content and presentation variables, and the relationship of these variables to student performance, thus providing a comprehensive definition of program structure and an evaluative model of program adaptations.

125

> The present computer program (which is now in its ninth revision) will take natural language input without editing from all of the following sources: computer assisted instructional terminal tapes, printer tapes, flex tapes or text cards; and convert the data to standard magnetic tape, according to a schema devised by the Rand Linguistic group. It will then produce a cross index of the materials, frequency counts of all variables and a prediction of the readability based on a regression equation. All of these are used to determine reading difficulty and program features.

These measures can be said to be easy to automate since they were designed to be calculated by hand and are based on the raw physical and linguistic characteristics of words and sentences. In contrast, the focus of the present report is on readability/comprehensibility measures which are characterized by their attempt to measure the intellective difficulty of the contents. Alternatively, we may say that the measures with which we deal have a goal of measuring difficulty of concepts--the amount of thinking which a reader will have to do. This, if you will, is the intellective work load that the reader must expend in order to gain an understanding of the meaning of the text. These measures, therefore, represent an attempt to quantify the complexity of what is happening inside a reader's head, rather than to determine comprehensibility purely on the basis of sizes and frequencies of words and sentences.

Comprehensibility Measures

The full 14 measures described in Chapters II and III will be discussed individually. It is assumed that the reader is familiar with the purpose and nature of these measures.

Each of the eight structure-of-intellect and six of the psycholinguistic measures associated with the readability/comprehensibility studies and listed in Table 4-1 will be presented in terms of a suggested approach toward computerization.

Throughout the discussion, it is assumed that the measures are calculated on a block of text whose size is variable, and that each variable is calculated for each text block.

Scaling will be such that higher values of the measures represent more difficult (less comprehensible) text and conversely lower values of the measures depict more readily understood writings.

Structure-of-Intellect Measures

The cognition of semantic units (CMU) measure was based on the type/token ratio. It seems that, in any given body of text, this factor can be readily automated by a series of word counts. A highly satisfactory value for this factor can be obtained through the straight-forward approach of calculating the ratio concerning a text block:

$$CMU = \frac{\text{Number of different words}}{\text{Total number of words}} = \frac{NDW(B)}{TNW(B)}$$

A number of decisions are inherent in this calculation. Principally any string of two or more symbols between two spaces will be tallied as a word. Preliminary analysis resolves these questions in the following manner:

(1) prefixes, tenses and the like will be taken into account (e.g., the word "walk" and "walking" will be tallied as two different words)

(2) abbreviations of multiple words, (e.g., "USAF," "USSR," "APA" will each be counted as one word). A count of the number of abbreviated words will be retained for use in calculating ESU below.

(3) hyphenated words will be counted as one word

(4) each word in a spelled out number will be counted as one word (e.g., "eight hundred" will be counted as two words)

(5) each numerical value (e.g., "485.6") will be counted as one word

(6) words printed in capital letter, italics, or foreign words will be tallied as individual words

(7) selected symbols will be contained in the dictionary (discussed below) and tallied as appropriate. Examples of word counts for sample symbols are:

symbol	number of words
&	1
¢	1
<	2
≤	5

This count will also be retained for use in calculating ESU below. Thus the calculation is considered to have no inherent technical risk.

The cognition of semantic relation (CMR) metric is defined as the number of sentences divided by the number of incomplete links or relations in a textual block. The former is, of course, much easier to determine than the latter. However, even the determination of the number of sentences in a given text is non-trivial. Its logic is discussed briefly by presenting the following considerations:

(1) codes will be used to identify portions which will not be involved in determination of this measure; e.g., tables, bibliographies, and figures will be bypassed

(2) sentences will be determined by scanning the periods, question marks, or exclamation marks designating the end of a sentence. The end of a sentence will be tallied only when one of these symbols follows a word or number (other than an abbreviation which will be checked against a prestored list) without an intervening space, and is followed by a space and a capital letter. (Both upper and lower case capability is assumed).

(3) logic will be required to distinguish an exclamation point from a factorial sign, and a period at the end of a sentence from one following an integer.

Determining the number of incomplete links or relations in text is substantially more difficult to automate and, as such, represents considerable technical risk. Detailed analyses would be required to obtain wholly a satisfactory logic and a resultant computer program. Such logic would involve identifying constructions such as compound subjects or predicates in conjunction with indefinite pronouns. It is anticipated that development of several limiting rules which define exceptions to general semantic relational logic would be a reasonable approach to this measure.

Memory of semantic units (MMU) is the next measure. It can be determined by a count of the number of fact repetitions per block of text. A simple approach is expected to yield satisfactory results. This measure would count words, phrases, and indicators which, in the English language, imply that a fact repetition is expected. Thus, MMU can be calculated by extension of the rules below:

(1) count one fact repetition for each occurrence of the following:

that is
i. e.
thus
consequently
in other words
therefore

(2) logic will be required to determine more precise conditions under which one fact repetition is counted for words such as the following:

repeat
accordingly
in effect
consequently
reiterate
recapitulate

(3) logic will be required to determine when two fact repetitions are counted when, under certain circumstances, one of these key words or phrases above is followed by "and"

Evaluation of symbolic implications is defined to be the ratio:

$$ESU = \frac{\text{number of abbreviated or symbolic words}}{\text{total number of words}} = \frac{NSW(B)}{TNW(B)}$$

Calculation of the ESU will be largely a byproduct of the determination of CMU and as such its calculation is considered relatively simple and risk free. The denominator of ESU and CMU are identified. The count of all abbreviations can be determined by the resultant count of multiple word abbreviations (from the CMU calculation) plus the count of single word abbreviations (e.g., Mr., Ave., and Pres.), plus the tallied word count results from symbols also determined in the calculation of CMU.

The value of ESU will therefore be scaled in the range 0-1 and in most cases is expected to assume low values, say below 0.1.

The cognition of figural units measure (CFU) is defined as the number of labelled locations or positions on a map, diagram, or drawing. A simple count of the number of textual (alpha-numeric) entries in a given diagram may be determined by a tally of such words or phrases as are provided as input to the text processing program. Here, however, we conceive only of textual input, not the graphics of a figure. Some logical rules for calculating CFU factor follow:

(1) a code will indicate the start and end of entries on the figures

(2) an independent programmatic check should be incorporated to identify all such words/phrases which are different

(3) name or indicators of several words will be counted as one

(4) the inclusion of abbreviations within a phrase will not alter the fact that the phrase will count as one

(5) abbreviations which stand alone (comprise a complete label entity) will also count as one

(6) each scale on a graph, title, column heading, figure name, map coordinate, and similar entity will count as one regardless of its size or number of characters

A preliminary version of the CMU measure can therefore be obtained, but the measure is unscaled and not comparable with the other measures. It generates data on each figure--not on the number of figures per block of text as do other measures. Therefore, initially it is recommended that CFU receive attention in the automatic determination of comprehensibility measures only in the generation of printed lists showing:

a. CFU for each figure
b. CFU per square inch of figure
c. mean and sigma of CFU per block

The next measure considered is convergent production of semantic systems (NMS). This measure may be defined as the number of mnemonic devices which are presented to the reader in a block of text (the art of strengthening the memory by using certain formal or mechanical methods of remembering is called mnemonics). Some examples of the use of mnemonic devices together with preliminary logic rules for their implementation are:

(1) the coining of a phrase or abbreviation in order to assist the reader in learning or remembering a concept. For example in the learning of the musical staff, the musician introduced to the mnemonic FACE as a way to remember the names of the notes between lines of the treble clef. As an extension of this mnemonic device, certain acronyms would qualify as mnemonic instances:

    FORTRAN - Formula Translation
    RADAR - Radio and Ranging

    In many cases the introduction of an abbreviation itself would qualify as an instance of mnemonic application. In these cases the mnemonics could be handled as dictionary entries (i.e., identified as a mnemonic device as part of the automated dictionary) and the first use of each would be tallied as part of the calculation of NMS.

(2) A mnemonic device can also take the form of an acrostic. For example, Psalm 145 is composed in such a way that the first letter of each line comprises the alphabet in sequence.

An acrostic not only could apply to first letters of words, but could also apply to middle or technical letters forming a word, or words or the regular or inverted alphabetical sequence. A jingle such as: "Thirty days hath September, etc." is also a valid example of a mnemonic. The identification of such cases, however simple and effective to the reader, can be most difficult to detect automatically in an efficient way. Additional work would be required in order to determine how to identify restrictions under which acrostics and jingles could.be counted.

(3) In some cases the display of a figure to describe a process or phenomenon would qualify as a mnemonic device. For example, if the explanation of the physical composition of the atom, relative to nucleus and orbiting electrons, was accompanied by a sketch which assisted the reader in understanding the concepts rather than in text alone, this would be tallied as a mnemonic device separately. The logic for this becomes complex due to the need to handle specific rather than general cases. However, it is recognized that not every figure, picture, or line drawing qualifies as a mnemonic device. The difference between a figure which serves as a memory assist and one which is presented merely to elaborate, to beautify, or to depict a scene is a very subtle one for which the success of automation is not obvious. This may require analyst precoding to separate. The logic for this becomes more complex due to the need to handle specific rather than general cases.

(4) As a last example of a mnemonic device, we cite the case of a symbolic formula, (e.g., $E = mc^2$). This is similar to an abbreviation in many cases, but symbols are substitutes for words, constants, variables, mathematical operations, and the like. Yet, here again, all formulas would not qualify to be tallied as mnemonic devices. Clearly, a proof or algebraic derivation involving n equalities stated symbolically would not qualify as n cases of mnemonic devices. Here again, more specific criteria as to precisely when to tally a specific case are required.

Accordingly, this variable demands considerable attention prior to implementation. This is due to the wide variety of types of mnemonic devices and their relative infrequency of occurrence. Relatively large expenditures of effort will be required to develop a variety of infrequently used logic which could add considerable complexity to the computer program.

The seventh structure-of-intellect derived readability/comprehensibility measure is the convergent production of semantic implications (NMI). This is defined as a tally of the number of times a synthesis of two or more items in the text is required but not provided.

The automatic determination of situations in which a synthesis of two elements is required in a body of text is an exceedingly difficult technical task. No known solution exists since the determination is tantamount to the requirement to determine whether or not a conclusion or a logical extention can be drawn from two (or more) sentences or phrases regardless of their placement within the text. Assuming this difficult determination, a somewhat less difficult problem would need solution, namely, an answer to the question: "Was this conclusion in fact drawn somewhere in the text?"

This clearly calls for techniques beyond the present scope of capabilities in language data processing, and further study would be required.

The last readability/comprehensibility measure described in this section is divergent production of semantic units (DMU). This measure is defined to equal the number of elucidations, explanations, or elucidations contained in the subject block of text. Presentation of an illustrative example in any form would meet this criteria. Here, further study will also be required to specify detailed implementation. However, the approach outlined above for MMU appears to provide a reasonable direction:

(1) Count one explanation for each occurrence of the following word or word combinations:

that is
i. e.
thus
consequently
in other words
therefore
to illustrate
for example

(2) logic will be required to determine specific conditions under which text including the following words or phrases is counted as one explanation:

elucidate
explain
illustrate
expound
instance
case
example

Psycholinguistic Measures

Yngve (1959) defined a new approach to measuring the depth or complexity of a sentence. This measure has come to be called Yngve depth. However, one who attempts to calculate the Yngve depth of a sentence will not find the exercise to be a relaxing way to pass his time. There has been published, however, (American Society for Information Sciences) a series of over one hundred sentence structure possibilities each with its precalculated Yngve depth value. The procedure recommended for implementation of the Yngve depth measure is one which will allow the computer to attempt to match each given sentence (in the text whose readability/comprehensibility is to be determined) to one of the available sentences with a precalculated depth value. This matching will be done on the basis of parts of speech, as follows: Each of the predetermined sentences will be manually parsed and the pattern sequence of parts of speech will be stored. The following basic parts of speech will be considered:

a-article
v-verb
adj- adjective
adv- adverb
n-noun
p-pronoun
c-conjunction
prep-preposition
e-exclamation

Accordingly, the sample sentence "The new club members came early" will be prestored as a sequence of parts of speech, and the depth value. Each sentence in the text to be measured will be parsed by the computer (either automatically or with the aid of some precoding) and compared against the prestored sentences, sorted in order by the number of words in the sentence. Therefore, the Yngve depth (YD) for the sentence under consideration will be that score as given with the prestored sentence which matches, or matches most closely, with the parts of speech sequence and the number of words.

Dealing with the problem of disambiguitization (i. e., in the example on the preceding page, whether the word "club" is a noun or a verb) represents the most difficult aspect of a completely satisfactory solution. Considerable technical risk is involved in achieving a full solution to this problem.

The second psycholinguistic measure is morpheme depth. The morpheme is "a linguistic or word unit which has no smaller meaningful parts." Alternately a morpheme is one or more syllables which together have some semantic meaning. For our purposes, the morpheme depth measure (MD) is determined by obtaining a tally of the number of morphemes in a block of text.

The best approach to automation of the morpheme depth is thought to be through a dictionary look-up procedure. To this end, it would be necessary to add to a currently available dictionary (in magnetic tape form) the number of morphemes corresponding to each dictionary entry. Thus, the word "unequivocal" which has five syllables (un-e-quiv-o-cal) would also be listed in the dictionary as having three morphemes (un-equi-vocal).

The tally of the morphemes would be accomplished using rules such as the following:

1. each numerical value (e. g., 3. 14159) will be counted as one morpheme

2. abbreviations whether one word (Mr.) or multiple words (USAF) will be talled as a single morpheme

3. capitalization will be required in morpheme counting

4. selected symbols will be included in the dictionary; for example, $\geq$ will be counted as one morpheme

5. since some morphemes are multiword (e. g., "for goodness sake") logic would be required to identify their occurrence from new dictionary entries

The calculation of the morpheme depth measure for any text block will be determined as the quotient of the morpheme count (total number of morphemes per block [TNM(B)], divided by the total number of words [TNW(B)], as calculated for CMU.

The next psycholinguistic measure to be calculated is transformational complexity (TC). It measures the number of transformations required to derive the "deep structure" from the "surface structure" of a sentence. The scoring here will be based on the count of the four types of sentences: (1) active, (2) active-negative, (3) passive, and (4) passive-negative. The basic problem here, then, is the definition of a logic suitable for identifying four categories of sentences. This is considered feasible within present capabilities.

A few of the characteristics of passive sentences are itemized below, with the understanding that a more complete logic may have to be devised prior to implementation on a computer. A sentence is passive when it contains:

1. two, three, or four verb words together or separated by one or two other words

2. the first of these verbs is one of the following forms of the verb to be:

is	be
is being	was to have been
was	will be
was being	will have been
has been	having been

3. the last of these verbs would be a past participle

4. for passive negative sentences, one of the following words or phrases must appear with the verbs mentioned above:

   not
   never
   n't

For active negative sentences, the computer would attempt to match a small selection of key negative words (in predetermined juxtapositions with respect to the sentence verb).

Additional ground rules would be programmed based on the dictionary lookup. Since many dictionary words will be categorized as either positive or negative and active or passive, this additional information will be utilized in making a selection as to the type of sentence. A more complete list of dictionary contents for each entry is given in Table 4-4.

Table 4-4

Dictionary Contents for Each Word

Contents	Admissible Values
positive/negative	P,N
active/passive	A,P
part of speech	up to 4 of 8 types
contained in Dales list of 3000	Y,N
contained in Dales list of 769	Y,N
no. of morphemes	1 thru 10
no. of syllables	1 thru 10
start of multi word morphemes	Y,N

Those sentences not categorized in the other three classes would be tallied as active. Thus, for each block, the total of active sentences per block [TAS(B)], and corresponding tallies of sentence types for passive [TPS(B)], active negative [TANS(B)], and passive negative [TPNS(B)] would be determined. Since the four categories of sentences have been shown to have differing levels of significance on comprehensibility, the four values representing the count of the number of each type of sentence will be multiplied by four weighting values submitted to the computer as run parameters. The weights will be representative of the level of significance on comprehensibility. Tentative value ranges for the weights are presented on the following page.

Item, I	Type of Sentence	Weight Range = WI
1	Active	1.0
2	Passive	1.0 - 1.1
3	Active Negative	1.5 - 2.0
4	Passive Negative	2.0 - 8.0

The final measure for transformational complexity, TC, would then be the scalar product of the four tallies, by the weights, divided by the number of sentences in the block:

$$TC(B) = \frac{TAS(B) \cdot W(1) + TPS(B) \cdot W(2) + TANS(B) \cdot W(3) + TPNS(B) \cdot W(4)}{NS(B)}$$

The fourth psycholinguistic measure of readability/comprehensibility is self embeddedness. One measure of embeddedness can be obtained by a tally of the number of words which separate the subject and the verb of the sentence. The problem here, as before, is automatic detection of the subject and verb in view of ambiguity of assignments of some words to parts of speech, particularly noun and verb interaction. However, assuming this problem to be solved for other measures, no additional parsing would be required for the self embedding measure (SE). The following illustrates the logic for this calculation:

(1) count words between the subject to the first verb. For a block of text, the total of such counts divided by the number of sentences is the self embedding measure.

(2) in case of sentences having more than one subject-verb pair, only the first pair will be counted.

141

(3) in cases of sentences having multiple subjects, the counting will begin with the last subject.

(4) the logic counting of abbreviations, symbols, and the like will be the same as that described for CMU above.

In the case of the sentence branching (SB) measure, we determine the placement of the verb in the sentence. The automation of this measure for any given sentence can be accomplished by the following:

(1) identify the principal word serving as the verb of the sentence

(2) count the number of words occurring in the sentence up to and including that verb, NWV

(3) count the number of words in the sentence, NWS

(4) calculate the ratio $\frac{NWV}{NWS}$. This is a number in the zero to one range indicating the placing of the verb.

For a block of text, the measure would be calculated as the average of all values obtained. The problem of identifying the verb has been discussed in prior sections. However, logic will be required for the compound sentence case.

The Processing Sequence

Figure 4-1 provides a preliminary sequential structure for computer calculating of the various readability/comprehensibility measures. A computer run will consist of processing one or more blocks of text, as a function of the text block size parameter, TBS. A value of TBS will specify the number of words in a text block. Comprehensibility measures will be calculated for each text block which equals or exceeds 100.

If TBS = 0, the computer program will scan for the code symbol sequences @@ and @@@. Each occurrence of @@ will signify the end of a block. In this way, the analyst can specify that measures be calculated for each section or chapter. The occurrence of @@@ will signify a request to summarize (averages to determine the measures for all text since the previous @@@ or since the start of text). This provides the ability to summarize over a volume having multiple sections or chapters. In Part I of Figure 4-1, a dictionary search is performed for all text in the block.

A tabulation of any word not found in the dictionary is presented to the analyst before the program can enter Part II. This protects against most spelling errors and improves the likelihood of valid processing later. In Part I, a magnetic tape will be prepared for each word of text based on the results of the dictionary lookup. Part I would be devised so that reruns can bypass the lookup except for new words.

In Part II, the process is performed sequentially on 100 word segments of text. In this part, the more mechanical tallies of words, syllables, etc., are performed. At the end of Part II, sufficient data will have been collected to calculate the "classical" values of reading grade level for the segments. Table 4-5 lists the various formulas extracted from the literature possible for calculation. Some or all of these will be incorporated into the program. Table 4-6 is a variable list for these and other variables.

Part I

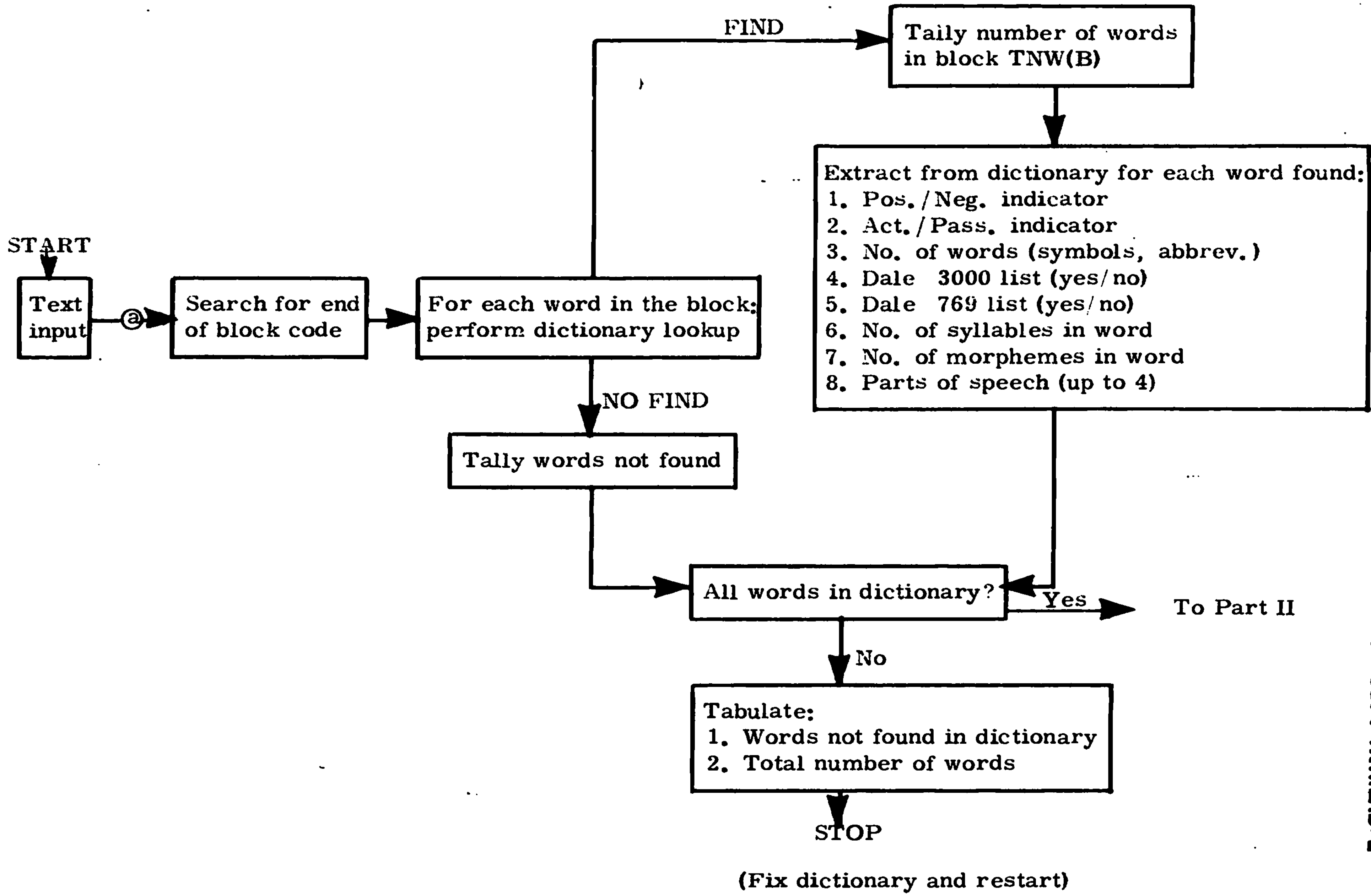

Figure 4-1. Logic flow chart for reading comprehendibility measures calculation.

144

Part II

Break text block into 100 word segments

Calculate for each segment, S = 1, ..., SMAX

1. No. of one syllable words, OSW(S)
2. No. of words not on Dale list of 769, AHW(S)
3. Average sentence length, ASL(S) = no. of words present
4. Average no. of prep. phrases, APP(S)
5. Average word length, AWL(S) = syllables per segment
6. No. of 2 syllable words, TSW(S)
7. No. of words not on Dale list of 3000, DALE
8. No. of 3 or more syllable words, MSW
9. Average no. of letters per word, ALW(S)
10. No. of sentences, NS(S)
11. No. of morphemes, TNM(S)

Calculate 9 values of reading grade level

To Part III

145

Part III

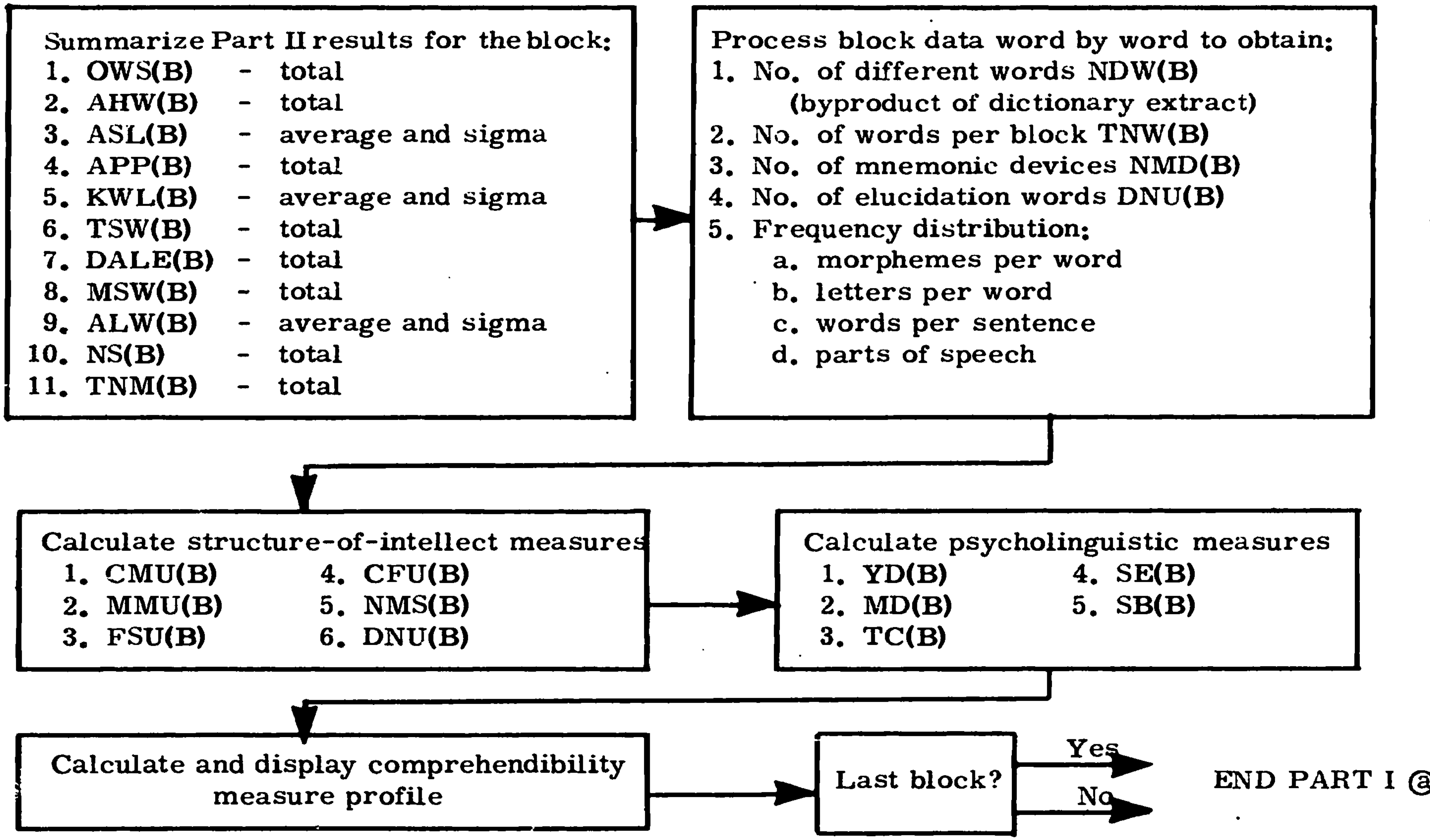
Summarize Part II results for the block:
1. OWS(B) - total
2. AHW(B) - total
3. ASL(B) - average and sigma
4. APP(B) - total
5. KWL(B) - average and sigma
6. TSW(B) - total
7. DALE(B) - total
8. MSW(B) - total
9. ALW(B) - average and sigma
10. NS(B) - total
11. TNM(B) - total
Process block data word by word to obtain:
1. No. of different words NDW(B)
(byproduct of dictionary extract)
2. No. of words per block TNW(B)
3. No. of mnemonic devices NMD(B)
4. No. of elucidation words DNU(B)
5. Frequency distribution:
a. morphemes per word
b. letters per word
c. words per sentence
d. parts of speech
Calculate structure-of-intellect measures
1. CMU(B)
2. MMU(B)
3. FSU(B)
4. CFU(B)
5. NMS(B)
6. DNU(B)
Calculate psycholinguistic measures
1. YD(B)
2. MD(B)
3. TC(B)
4. SE(B)
5. SB(B)
Calculate and display comprehendibility measure profile
Last block?
Yes
No
END PART I ⓐ

Table 4-5

Reading Grade Level Formulas from the Literature

	Formula	Author
	$RGL = 20 - \frac{OSW}{10}$	Caylor/Sticht/Fox/Ford
	$RGL = 0.5\ ASL + 4.71\ ALW - 21.43$	Smith/Senter
50% Level:	$RGL = 0.1\ AHW + 0.60\ ASL + 0.1\ APP + 1.99$	Lorge/Dale
50%	$RGL = 0.0778\ ASL + 0.0455\ AWL - 2.2029$	Flesch
100%	$RGL = 3 + \sqrt{MSWL}$	McLaughlin
	$RGL = 0.0596\ ASL + 0.1155\ (DALE) + 3.2672$	Dale/Chall
	$RGL = 0.0923\ ASL - 0.0648\ AWL + 8.4355$	Farr/Jenkins/Patterson
	$RGL = 0.0984\ MSW$	Gun·ing
	$RGL = \begin{cases} \frac{OSW + TSW + 3\ MSW}{2} & \text{it } OSW + TSW + 3MSW > 20 \\ \frac{OSW + TSW + 3\ MSW - 2}{2} & \text{otherwise} \end{cases}$	McElroy

BEST COPY AVAILABLE

147

Table 4-6

Initial List of Variables

Fortran Code	Variable
S	Segment of 100 words
B	Block of up to 100 words
NDW(B)	Number of different words in a block
TNW(B)	Total number of words in a block
NS(B)	Number of sentences in a block
NSW(B)	Number of symbolic words in a block
NMD(B)	Number of mnemonic devices in a block
TNM(B)	Total number of morphemes in a block
TAS(B)	Total number of active sentences in a block
TPS(B)	Total number of passive sentences in a block
TANS(B)	Total number of active negative sentences in a block
TPNS(B)	Total number of negative passive sentences in a block
CMU(B)	Cognition of semantic units (type token ratio)
CMR(B)	Cognition of semantic relations
ESU(B)	Evaluation of symbolic implications
CFU(B)	Cognition of figural units
NMS(B)	Number of mnemonic systems (count of mnemonics)
NMI(B)	Convergent production of semantic implications
DNU(B)	Divergent production of semantic units
YD	Yngve depth measure
MD(B)	Morpheme depth measure
TC(B)	Transformational complexity measure
SE(B)	self embedding measure
TBS(B)	Text block size
RGL	Reading grade level
AHW	Average no. of hard words (words not in Dale's list of 769 entry words) per 100 word sample
ASL	Average sentence length in words
APP	Average no. of prepositional phrases per 100 words
AWL	Average word length= number of syllables per 100 words
DSW	Number of one syllable words per 100 words
TSW	Number of two syllable words per 100 words
DALE	Dale score, the numbers of words per 100 words not appearing in list of 3,000 words known to 80% sample of 4th graders
MSW	No. of words of 3 or more syllables per 100 words
ALW	Average strokes (letters) per word= word length
MSWL	MSW per 30 sentences

Part III processing, on a block size basis, starts with the summarization of Part II data over the block. Part III includes the calculation of the readability/comprehensibility measures as described in this chapter.

The result of Part III is a comprehensibility profile. This would take the form of a listing of all of the comprehensibility measures. In addition, these would be further processed by weights, scaling adjustments, and algebraic combination into one or two comprehensibility indices for the particular data block. Processing continues for each text block. Thus, the total outputs include: averages, frequency distributions, and final indices for each block. The capability of summarizing over block results, to effectively record volume results from the sum of its chapters, would also be provided.

CHAPTER V

FINAL WORDS

What then can be said, in summary, about the role of structure-of-intellect oriented and psycholinguistically based variables vis-a-vis the issue of measuring and increasing the readability/comprehensibility of reading materials? The results presented in Chapter II clearly support a contention that measurement of the intellective load imposed by textual material on the reader, through structure-of-intellect based variables, will tell us something about the readability/comprehensibility of the text. Additionally, the psycholinguistic investigations reported in Chapter III yielded a set of results which substantiated the value (for the most part) of the psycholinguistic pathway.

Admittedly, we do not know whether or not the two approaches are truly independent. For example, it seems quite probable that the memory for semantic relations structure-of-intellect concept in the comprehensibility sphere is analogous (i.e., based on the same ability) to the left-right branching psycholinguistic concept. Similarly, the morpheme volume and memory for semantic unit variables may be related. Description of the same phenomenon in different terms does not represent an alien situation. This holds for both the behavioral sciences (e.g., learning theory or personality theory) and the physical sciences (e.g., nerve impulse transmission or electron flow). On the other hand, different materials were employed in the two investigations. Accordingly, there is no way of knowing, from the present work, the degree of correlation among the various concepts involved.

A similar question is concerned with the relationship between the structure-of-intellect variables and the psycholinguistic variables on the one hand, and prior measures of readability/comprehensibility on the other hand. To provide some measures of this relationship, the structure-of-intellect stimulus materials were subjected to Flesch analysis and to automated readability index (ARI) analysis. The obtained Flesch and ARI scores were then correlated with the scores of the materials on the structure-of-intellect measures. The results (phi coefficients) indicate a rather large degree of independence of the structure-of-intellect oriented comprehensibility analysis from these two prior techniques. While similar data are not as yet developed for the psycholinguistic data, there is little reason to believe that a similar result would not obtain. For example, the Flesch or the ARI techniques (which count words) would not discriminate between two matched sentences, one of which is right branching, while the other is left branching. The psycholinguistic measure does.

Moreover, the present set of techniques will tell the user not only that a given set of material is more or less readable/comprehensible than another text, but also what steps should be taken to increase the readability/comprehensibility. Accordingly, the new techniques possess diagnostic as well as interpretive value. This is not true for prior techniques. In fact, Flesch warned that his technique is not to be used to develop rules for writing readable text. He advised use of his technique only for measuring readability. On the other hand, the structure-of-intellect and the psycholinguistically based concepts provide a basis for writing text which will be readable/comprehensible. Currently, a procedural guide is under development which will state how these variables can be measured by interested users. It is anticipated that these procedures will be of considerable interest to persons who prepare Air Force training materials.

To the degree that the required measurements can be made by others, the techniques here developed can be held to be useful. And, utility is considered to be one criterion for judging the merit of any new technique. Related to the problem of technique utility is application ease. Presently, the structure-of-intellect and the psycholinguistic measures rest on hand calculations--as is true for any of the other readability measures, with the exception of the ARI. However, chapter IV of the present report describes the potential for automating the determination of a large number of these variables.

Other criteria for judging the merit of any new technique rest on considerations of psychometric reliability and validity. There is little, if any, reason to suspect that the within or the between user reliabilities of the present techniques are unacceptably low. Both techniques are based on objective counts and the like. These counts can be defined and methods for their derivation can be concretely specified. Accordingly, users who can be taught to follow concrete rules should obtain acceptable reliability in the use of the techniques. These arguments, however, do not obviate the need for studies into the reliability of the techniques in actual application.

Contentions supporting the validity of the new techniques must rest on arguments relative to their construct validity and their predictive validity. Construct validity is evaluated on the basis of the psychological qualities that a technique measures. Quite obviously, the thrust of the expository aspects of the present report was oriented toward arguments supporting the construct validity of the structure-of-intellect and the psycholinguistic variables in the readability/comprehensibility context. Predictive validity is evaluated by showing how well predictions of a technique are confirmed by evidence collected at some subsequent time. Equally obvious is the thrust of the research reported in chapters III and IV, which focused on the establishment of the predictive validity of the various measures. The research results substantiate a contention of predictive validity for a large number of the variables investigated. Cross validation of any set of findings is always warranted. Certainly, it is warranted here in view of: (1) the potential of the present findings for achieving a major contribution in increasing the abilityof written materials to transmit information, and (2) the novelty of the concepts presented.

Finally, the present set of studies was concerned only with a subset of structure-of-intellect and psycholinguistic variables. Those variables which seemed most relevant to our purposes, those which were most easily quantified, and those which seemed most objective were selected for this initial investigation. The potential of other variables, both psycholinguistic and structure-of-intellect, should be investigated in the readability/comprehensibility context. There is little pedagological value in making the reader work hard to benefit from the written word. The written word is with us and will stay with us in the foreseeable future. One would not produce a book from a print that is blurry. Why must concepts be presented in a blurred manner?

REFERENCES

Baer, R. & Odbert, H. Insight of older pupils into their knowledge and word meanings. *School Review*, 1941, *49*, 754-760.

Berelson, B. Content analysis. In G. Lindsey (Ed.). *Handbook of social psychology.* Cambridge: Harvard University Press, 1954.

Bobbitt, J. An experimental study of the phenomenon of closure as a threshold function. In D. Breadslee & M. Wertheimer, *Readings in perception.* Princeton: Von Nostrand, 1958, 136-159.

Borko, H. *Computer applications in the behavioral sciences.* Englewood Cliff: Prentice-Hall, 1962.

Bormuth, J.R. *Development of readability.* Final Report, U.S. Department of Health, Education, and Welfare, Project No. 7-0052, 1969.

Bowles, E.A. *Computers in humanistic research.* Englewood Cliff: Prentice-Hall, 1967.

Bourisseau, W., Davis, O., & Yamamato, K. Semi-impression responses to differing pictorial and verbal stimuli. *AV Communication Review*, 1965, *13*, 249-258.

Briggs, L. A procedure for the design of multimedia instruction. In: *Proceedings of the 74th Annual Convention of the American Psychological Association.* Washington, D.C.: American Psychological Association, 1966, Pp. 257-258.

Caylor, J.S., Sticht, T.G., Fox, L.C., & Ford, J.P. *Metgodologies for determining reading requirements of military occupational specialties.* Presidio of Monterey, Calif.: Human Resources Research Organization, HumRRO Division No. 3, Technical Report 73-5, March 1973.

Chapanis, A. Words, words, words. *Human Factors*, 1965, *7*, 1-17.

Chomsky, N. *Syntactic structures.* The Hague: Mouton, 1957.

Clark, H.H., & Chase, W.G. On the process of comparing sentences against pictures. *Journal of Cognitive Psychology,* 1972, *3*, 472-517.

Coke, E.V., & Rothkopf, E.Z. Bote of a simple algorith, for a computer produced reading ease score. *Journal of Applied Psychology,* 1970, *54*, 208-210.

Coleman, E.B. The comprehensibility of seven grammatical transformations. *Journal of Applied Psychology,* 1964, *48*, 186-190.

Coleman, E.B. Learning of prose written in four grammatical transformations. *Journal of Applied Psychology,* 1965, *49*, 332-342.

Coleman, E.B. Developing a technology of written instruction: Some determiners of the complexity of prose. In E.Z. Rothkopf & P.E. Johnson *Verbal learning research and the technology of written instruction.* New York: Teachers College Press, 1971, 155-204.

Combs, A., & Snygg, D. *Individual behavior.* New York: Harper, 1959.

Dale, E. A comparison of two word lists. *Educational Research Bulletin,* 1931, *10*, 484-489.

Dale, E., & Tyler, R. A study of the factors influencing the difficulty of reading materials for adults of limited reading ability. *Library Quarterly,* 1934, *4*, 384-412.

Dale, E., & Chall, J.S. A formula for predicting readability (and) instructions. *Educational Research Bulletin,* 1948, *27*, 11-20, 28, 37-54.

Eiferman, R.R. Negation: A linguistic variable. *Acta Psychologica,* 1961, *18*, 250-273.

Faden, B.R. *Computer programs dictionary.* New York: Crowell, Collier, & MacMillan, 1971.

Flesch, R. A new readability yardstick. *Journal of Applied Psychology*, 1948, *32*, 221-233.

Flesch, R. Estimating the comprehension difficulty of magazine articles. *Journal of General Psychology*, 1943, *28*, 63-80.

Fodor, J., & Garrett, M. Some syntactic determinants of sential complexity. *Perception and Psychophysics*, 1967, *2*, 289-296.

Fodor, J.A. Current approaches to syntax recognition. In Horton, D.L., & Jenkins, J.J. (Eds.), *Perception of language*, Columbus, Ohio: Charles E. Merrill Publishing Company, 1971.

Foss, D.J., & Crains, H.S. Some effects of memory limitation upon sentence comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 541-547.

Francis, W.N. *A standard corpus of edited present day American English for computer use.* Proceedings of the IBM Literary Data Processing Conference, 1964, 179-189.

Francis, W.N. *Manual of information to accompany a standard sample of present-day edited American English for use with digital computers.* Providence: Brown University, 1971.

Fredricksen, C.H. Semantic information processing in the acquisition of knowledge from written text. Paper presented at symposium on *Methodologies for research on written materials*, American Psychological Association, Montreal, August 1973.

Gagné, R. *The conditions of learning.* Holt, Rinehart, & Winston, 1965.

Gough, P.B. Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior*, 1965, *4*, 107-111.

Gray, W., & Leary, B. *What makes a book readable.* Chicago: University of Chicago Press, 1935.

Green, B.F. *Digital computers in research. An introduction for behavioral and social scientists.* New York: McGraw-Hill, 1969.

Hamilton, H.W., & Deese, J. Comprehensibility and subject-verb relations in complex sentences. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 163-170.

Hakes, D.T. Effects reducing complement constructions on sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 278-286.

Harris, A. *How to increase reading ability: A guide to developmental and remedial methods.* New York: McKay, 1961.

Hilgard, E. *Theories of learning.* New York: Appleton-Century Crofts, 1956.

Honseholder, F.W., Jr. The information of general semantics. Quarterly Report 3. Contract AF-3 (602) 2184. Indiana University NTIS Report AD-250-449 13 Feb. 1961. American Society for Information Sciences. Document 008292, New York: CCM Information Sciences, undated.

Hovland, C. Human learning and retention. In S.S. Stevens (Ed.) *Handbook of Experimental Psychology.* New York: Wiley, 1951.

Jacobson, M.D., & MacDougall, M.A. *Computerized model of program structure and learning difficulty.* Proceedings of the American Psychological Association. 77th. Annual Convention, 1969.

Jacobson, M.D., & MacDougall, M.A. Computer management of information and structure in computer supported instruction material. *Educational Technology*, 1970.

Johnson, D.O. Computer frequency control of vocabluary in language learning reading material. *Instructional Science.* Hseier Publishing Company, Amsterdam, Netherlands. p 121-131.

Lautman, M.R., Siegel, A.I., Williams, A.R., & Burkett, J.R. *The effects of reading difficulty, literacy level, and method of presentation on the comprehensibility of technical information training manuals.* Air Force Human Resources Laboratory, Technical Report, Lowry Air Force Base, Colorado, (in press).

Lindquist, E.J. *Design and analysis of experiements in psychology and education.* Boston: Houghton Mifflin, 1953.

Lorge, I. Predicting readability. *Teachers College Record*, 1944, *45*, 404-409.

Luhn, H.P. *Keyword in context index for technical literature.* (KWIC Index): IBM Corporation: ASDD Report R C-127, 1959.

Lumsdaine, A. Educational technology, programmed learning, and instructional science. *Theories of learning and instruction.* 63rd Year Book, National Society for the Study of Education, Chicago: University of Chicago Press, 1964, 371-401.

Madden, H.L., & Tupes, E.C. *Estimating reading ability level from the AQE General Aptitude Index.* Lackland Air Force Base, Texas: Personnel Research Laboratory, Aerospace Medical Division, PRL-TR-66-1, February, 1966. (AD-632 182)

Martin, E., & Roberts, K.H. Grammatical factors in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 211-218.

McMahon, L. *Grammatical analysis as part of understanding a sentence.* Unpublished doctoral dissertation, Harvard University, 1963.

Mehler, J. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 346-351.

Miller, G. Speech and language. In S.S. Stevens (Ed.), *Handbook of Experimental Psychology*, New York: Wiley, 1951.

Miller, G.A. The magical number seven, plus or minus two. *Psychological Review*, 1956, *63*, 81-97.

Miller, G.A. The magical number seven, plus or minus two: Score limi' on our capacity for processing information. In D. Beardsly & M. Wertheimer (Eds.). *Readings in perception.* Princeton: Von Nostrand, 1958, 90-114.

Miller, G.A. Some psychological studies of grammar. *American Psychologist*, 1962, *17*, 748-762.

Minsky, M. *Semantic information processing.* Cambridge, Mass.: MIT Press, 1969.

Moore, J.C., & Biederman, I. *Speeded recognition of ungrammaticality: Multiple violations.* Paper presented at the meeting of the Psychonomic Society, St. Louis, November 1973.

Mosteller, F., & Wallace, D. Inference in an authorship problem. *Journal of the American Statistical Association*, 1963, *58*, 275-309.

Ojemann, R. The reading ability of parents and factors associated with the reading difficulty of parent education materials. In G.D. Stoddard (Ed.) *Researches in parent education II.* Iowa City, Ia.: University of Iowa, 1934. Cited by J.S. Chall, *Readability-an appraisal of research and application.* Columbus, Ohio: Ohio State University, Bureau of Educational Research, 1958.

Olney, J., & Ramsey, D. From machine-readable dictionaries to a lexicon tester: Progress, plans, and an offer. *Computer studies in the humanistic and verbal behavior*, November 1972.

Osgood, C. *Method and theory in experimental psychology.* New York: Oxford, 1953.

Parrish, S.M. *Concordances to the poems of Mathew Arnold.* Itheca: Cornell University Press, 1959.

Perfetti, C.A. Sentence retention and the depth hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 101-104.

Pettinger, O., & Gooding, C. *Learning theories in educational practice: An integration of psychological theory and educational philosophy.* New York: Wiley, 1971.

Reichert, R., Olney, J., & Paris, J. *Two dictionary transcripts and programs for processing them.* Santa Monica: Systems Development Corporation, TM 3978, 1969.

Roberts, K.H. Lexial density and phrase structure depth as variables in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 719-724.

Rohrman, N.L. The role of syntactic structure in the recall of the nominalizations. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 904-912.

Ruben, E. Figure and ground. In D. Beardslee & M. Wertheimer (Eds.) *Readings in perception.* Princeton: Von Nostrand, 1958, 194-203.

Schwartz, D., Sparkman, J.P., & Deese, J. The process of understanding and jusgments of comprehensibility. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 87-93.

Sedelow, S. The computer in the humanistic and fine arts. *Association for Computing Machinery Computing Surveys*, 1970, *2*, 89-110.

Sherman, M.A. Bound to be easier? The negative prefix and sentence comprehension. *Journal of Verbal Learning and Behavior*, 1973, *12*, 78-84.

Siegel, A.I., & Siegel, E. Flesch readability analysis of the major preelection speeches of Eisenhower and Stevenson. *Journal of Applied Psychology*, 1953, *37*, 105-106.

Siegel S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Slobin, D.I. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 219-227.

Smith, E.A., & Senter, R.J. *Automated readability index.* Aerospace Medical Research Laboratories Technical Report 66-220. Wright-Patterson Air Force Base, Ohio. November 1970.

Spache, G. A new readability formula for primary grade reading materials. *Elementary school Journal*, 1953, *53*, 410-413.

Staats, A., & Staats, C. *Complex human behavior: A systematic extension of learning principles.* New York: Holt, Rinehart, & Winston, 1963.

Trabasso, T., Rollins, H., & Shaughnessy, E. Storage and verification stage in processing concepts. *Cognitive Psychology*, 1971, *2*, 239-289.

Trabasso, T. *Reasoning and the processing of negative information.* Invited address to Division 3 of the American Psychological Association, Miami, September 1973.

Underwood, B. *Experimental psychology.* New York: Appleton-Century Crofts, 1949.

Van Dorn, A., & Rice, D.E. On-line text editing: A survey. *Association for Computing Machinery Surveys,* 1971, *3,* 93-114.

Wang, M.D. The role of syntactic complexity as a determiner of comprehensibility. *Journal of Verbal Learning and Verbal Behavior,* 1970, *9,* 398-404.

Wason, P.S. The processing of a positive and negative information. *Quarterly Journal of Experimental Psychology,* 1959, *11,* 92-107.

Wason, P.C. Response to affirmative and negative binary statements. *British Journal of Psychology,* 1961, *52,* 133-142.

Wertheimer, M. Principles of Perceptual organization. In D. Beardslee & M. Wertheimer (Eds.) *Readings in perception.* Princeton: Von Nostrand, 1958, 115-135.

Williams, A.R., Jr., Siegel, A.I., & Burkett, J.R. *Readability of textual material--A survey of the literature.* **AFHRL-TR-74-29. Lowry AFB, Colo.: Technical Training Division, Air Force Human Resources Laboratory, July 1974.**

Woodworth, R., & Schlossberg, H. *Experimental psychology.* New York: Holt, 1954.

Wright, P. Two studies of the depth hypothesis. *British Journal of Psychology,* 1969, *60,* 63-69.

Yngve, V.H. A model and a hypotheses for language structure. Proceedings of the American Philosophical Society, 1960, *104,* 444-466.

Yngve, V.H. Implications of mechanical translation research. Proceedings of the American Philosophical Society, 1964, *108,* 275-281.